



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV RADIOELEKTRONIKY

DEPARTMENT OF RADIO ELECTRONICS

ANALÝZA HLASIVKOVÝCH PULZŮ

ANALYSIS OF GLOTTAL PULSES

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. David Příkazský

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. Milan Sigmund, CSc.

BRNO 2018

Diplomová práce

magisterský navazující studijní obor **Elektronika a sdělovací technika**

Ústav radioelektroniky

Student: Bc. David Příkazský

ID: 164888

Ročník: 2

Akademický rok: 2017/18

NÁZEV TÉMATU:

Analýza hlasivkových pulzů

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s problematikou hlasivkových pulzů. Vytvořte přehledovou studii o nejčastějších metodách na přístrojové měření tvaru hlasivkových pulzů. Vytvořte přehled softwarových nástrojů k určování hlasivkových pulzů z řečových signálů. Naprogramujte vybrané algoritmy na získávání hlasivkových pulzů z řečových záznamů a porovnejte získané výsledky s výstupy softwaru Aparat. Vytvořte ucelený program v prostředí Matlab GUI zahrnující jednotlivé algoritmy na určování hlasivkových pulzů a jejich spekter. Výstupem by měly být údaje o pulzech v grafické a číselné podobě jak pro krátké úseky řeči (časové průběhy, spektrogramy), tak pro dlouhé úseky řeči (histogramy).

DOPORUČENÁ LITERATURA:

[1] PSUTKA, J., MÜLLER, Z., MATOUŠEK, J., RADOVÁ, V. Mluvíme s počítačem česky. Praha: Academia, 2006.

[2] AIRAS, M., PULAKKA, H., BÄCKSTRÖM, T., ALKU, P. A toolkit for voice inverse filtering and parametrisation. In Proceedings INTERSPEECH 2005. Lisbon: 2005, p. 2145-2148.

Termín zadání: 5. 2. 2018

Termín odevzdání: 17. 5. 2018

Vedoucí práce: prof. Ing. Milan Sigmund, CSc.

prof. Ing. Tomáš Kratochvíl, Ph.D.
předseda oborové rady



UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

ABSTRAKT

Práce se zabývá odhadem hlasivkových pulzů z řečového záznamu. Je zde popsán proces tvorby řeči, dále popis přístrojů pro měření hlasivkových pulzů, přehled softwarových nástrojů umožňující odhad hlasivkových pulzů z řečového signálu. Popis metody IAIF a Sahoo –vy metody pro odhad hlasivkových pulzů. Pro snadnější ovládání zmíněných metod je vytvořeno grafické uživatelské prostředí (GUI) v programu MATLAB.

KLÍČOVÁ SLOVA

Hlasivkové pulzy, měření hlasivkových pulzů, odhad hlasivkových pulzů, tvorba mluvené řeči, LF-model, grafické uživatelské prostředí, GUI, MATLAB.

ABSTRACT

The work is about the estimation of vocal pulses from the speech record. Contains a description of the process of speech production, description of the instruments for the measurement of vocal pulses, an overview of software tools for estimating vocal pulses from the speech signal. Description of IAIF and Sahoo method for estimating vocal pulses. The Graphic User Interface in MATLAB is created for easier control of mentioned methods.

KEYWORDS

Glottal pulses, glottal flow, measurement of glottal pulses, estimate of glottal pulses, creation of spoken speech, LF-model, Graphic User Interface, GUI, MATLAB.

PŘÍKAZSKÝ, D. *Analýza hlasivkových pulzů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018. 63 s. Vedoucí diplomové práce prof. Ing. Milan Sigmund, CSc..

PROHLÁŠENÍ

Prohlašuji, že svoji diplomovou práci na téma Analýza hlasivkových pulzů jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne

.....

(podpis autora)

PODĚKOVÁNÍ

Děkuji vedoucímu diplomové práce prof. Ing. Milan Sigmund, CSc., za účinnou metodickou, pedagogickou i odbornou pomoc a další cenné rady při zpracování mé diplomové práce.

V Brně dne

.....

(podpis autora)

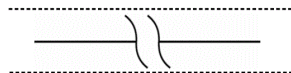
Experimentální část této diplomové práce byla realizována na výzkumné infrastruktuře
vybudované v rámci projektu CZ.1.05/2.1.00/03.0072
Centrum senzorických, informačních a komunikačních systémů (SIX)
operačního programu Výzkum a vývoj pro inovace.

Seznam použitých symbolů a zkratk	x
Seznam obrázků	xi
Seznam tabulek	xiii
Úvod	1
1 Vznik a vlastnosti mluvené řeči	2
1.1 Proces vytváření řeči člověkem	2
1.1.1 Dechové ústrojí	2
1.1.2 Hlasové ústrojí	3
1.1.3 Artikulační ústrojí	5
1.2 Přístrojové měření hlasivkových pulzů.....	5
1.2.1 Elektrolottograf (EGG), laryngograf	6
1.2.2 Elektromagnetický glottograf (EMGG).....	6
1.2.3 Měřicí sestava „Aeroview System“	7
1.2.4 Videolaryngostroboskopie	7
1.2.5 Videokymografie	8
1.2.6 Vysokorychlostní videolaryngoskopie	9
1.3 Akustické (programové) měření hlasivkových pulzů.....	9
1.3.1 Aalto Aparat.....	9
1.3.2 Sopran (DeCap)	10
1.3.3 Inverse filtering (UCLA Bureau of Glottal Affairs).....	11
1.4 Využití hlasivkových pulzů	11
1.4.1 Řečová syntéza	12
1.4.2 Rozpoznání mluvčího	13
1.4.3 Medicínské využití.....	14
2 Získání hlasivkových pulzů ze záznamu řečového signálu	16
2.1 Metoda iterativně adaptivního inverzního filtrování (Iterative Adaptive Inverse Filtering (IAIF)).....	16
2.2 Sahoo -va metoda pro inverzní filtrování	19
2.2.1 Stavové rovnice popisující modelu hlasového traktu	19
2.2.2 Lineární prostorový stavový model	20
2.2.3 Stavové rovnice z modelu hlasivkového pulzu.....	23
2.2.4 Nelineární prostorový stavový model	24

2.2.5	Realizace rozšířeného Kalmanova filtrování (EKF).....	26
2.2.6	Metoda odhadu hlasivkového pulzu	28
3	Odhad tvarů hlasivkových pulzů	29
3.1	IAIF metoda.....	29
3.2	Sahoo –va metoda.....	30
4	Popis grafického uživatelského prostředí pro snadné získání hlasivkových pulzů	32
4.1	Ukázka vytvořeného GUI	32
4.1.1	GUI, 1. blok: načtení řečového signálu a ukázání jeho časového průběhu	34
4.1.2	GUI, 2. blok: výběr části signálu jaká má být analyzována	34
4.1.3	GUI, 3. blok: výběr výpočetní metody pro získání hlasivkových pulzů	37
4.1.4	GUI, 4. blok I.: nastavení základních hodnot pro metodu IAIF	37
4.1.5	GUI, 4. blok II.: nastavení základních hodnot pro metodu Sahoo.....	38
4.1.6	GUI, 5. blok I.: zahájení výpočtu pomocí zvolené a nakonfigurované metody	39
4.1.7	GUI, 5. blok II.: výpočet hlasivkových pulzů při volbě „ <i>Every part with new setting</i> “ z 2. bloku	39
4.1.8	GUI, blok: graf „Calculated glottal flow“	40
4.1.9	GUI, blok: graf „Glottal flow data ready on export“	41
4.1.10	GUI, blok: Find glottal pulses.....	42
4.1.11	GUI, blok: Show (compare) glottal flow	44
4.1.12	GUI, blok: Export glottal flow data to	44
4.2	Podrobnější popis určitých funkcí pro bloky GUI.....	45
4.2.1	GUI, blok: graf „Calculated glottal flow“	45
4.2.2	GUI, blok: Find glottal pulses.....	45
4.2.3	GUI, blok: Show (compare) glottal flow	48
4.2.4	GUI, blok: Export glottal flow data to	49
5	Ukázka využití tvaru získaných hlasivkových pulzů	50
	Závěr	53
	Literatura	54
	Seznam příloh	57

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

Symboly:



... část obrázku mezi čárkováním byla vyříznuta z důvodu zmenšení obrázku

F_0

... frekvence základního hlasivkového tónu

EGG

... elektroglograf

IAIF

... Iterative Adaptive Inverse Filtering

LF

... Liljencrants-Fant (model)

LPC

... linear predictive coding

DAP

... discrete all-pole modeling

EKF

... extended Kalman filtering

GCI

... glottal closure instant

GOI

... glottal opening instant

GUI

... graphic user interface

SEZNAM OBRÁZKŮ

Obrázek 1.1	Schématický model hlasového ústrojí člověka [1].	2
Obrázek 1.2	Hlasový trakt člověka [1].	3
Obrázek 1.3	Hrtan a hlasivky (pohled shora) [1].	4
Obrázek 1.4	Ukázka tvaru hlasivkové štěrbině při klidu a) a fonačním postavením b) [1].	4
Obrázek 1.5	Hlasivkový signál změřený elektrolottografem [1].	5
Obrázek 1.6	Ukázka přístroje elektrolottografu s výstupními daty a umístěním elektrod na krku [2].	6
Obrázek 1.7	Umístění antén EMGG a elektrod EGG na krku [3].	6
Obrázek 1.8	Vlevo měřicí sestava „Aeroview System“, vpravo program Waveview [4].	7
Obrázek 1.9	Endoskop pevného (rigidního) tvaru [6].	7
Obrázek 1.10	Endoskop ohebného (flexibilního) tvaru [6].	8
Obrázek 1.11	Vlevo hlasivky s vyznačeným snímaným řádkem, vpravo výstup metody [8].	8
Obrázek 1.12	Ukázka prostředí programu Aalto Aparat [9].	10
Obrázek 1.13	Ukázka inverzního filtrování v programu Sopran [10].	10
Obrázek 1.14	Ukázka inverzního filtrování v programu Inverse filtering (UCLA Bureau of Glottal Affairs) [11].	11
Obrázek 1.15	Hlasivkový pulz podle Liljencrants-Fant modelu [17].	12
Obrázek 2.1	Model tvorby řečového signálu.	16
Obrázek 2.2	Vývojový diagram algoritmu IAIF.	18
Obrázek 2.3	Válcový model hlasového traktu [30].	19
Obrázek 2.4	Derivovaný hlasivkový pulz LF-modelu s označením důležitých okamžiků [24].	23
Obrázek 3.1	Odhadnuté hlasivkové pulzy samohlásky /a/, metoda IAIF. Pro názornost byl na řečový signál aplikován offset.	29
Obrázek 3.2	Odhadnuté hlasivkové pulzy slova „OKA“, metoda IAIF. Pro názornost byl na řečový signál aplikován offset.	30
Obrázek 3.3	Odhadnuté hlasivkové pulzy samohlásky /a/, Sahoo –va metoda. Pro názornost byl na řečový signál aplikován offset.	31
Obrázek 3.4	Odhadnuté hlasivkové pulzy slova „OKA“, Sahoo –va metoda. Pro názornost byl na řečový signál aplikován offset.	31

Obrázek 4.1	Vytvořené grafické uživatelské prostředí pro snadné získání hlasivkových pulzů při zpracovávání krátké zvukové nahrávky.	33
Obrázek 4.2	GUI, 1. blok: načtení řečového signálu a ukázání jeho časového průběhu.	34
Obrázek 4.3	GUI, 2. blok: část pro výběr celého signálu.	35
Obrázek 4.4	GUI, 2. blok: část pro výběr požadovaného časového úseku z celého signálu.	35
Obrázek 4.5	GUI, 2. blok: detekování znělých částí řeči z řečového signálu.	36
Obrázek 4.6	GUI, 3. blok: výběr výpočetní metody pro získání hlasivkových pulzů.	37
Obrázek 4.7	GUI, 4. blok I.: nastavení základních hodnot pro metodu IAIF.	38
Obrázek 4.8	GUI, 4. blok II.: nastavení základních hodnot pro metodu Sahoo.	38
Obrázek 4.9	GUI, 5. blok I.: a) blok obstarávající spuštění výpočtu pomocí zvolené a nakonfigurované metody; b) zpráva zobrazená při výpočtu metodou IAIF; c) zpráva zobrazená při výpočtu metodou Sahoo.	39
Obrázek 4.10	GUI, 5. blok II.: blok pro výpočet hlasivkových pulzů při volbě „ <i>Every part with new setting</i> “ z 2. bloku.	40
Obrázek 4.11	GUI, blok: graf „ Calculated glottal flow “ jeho význam a ovládací prvky.	41
Obrázek 4.12	GUI, blok: graf „ Glottal flow data ready on export “ jeho význam a ovládací prvky.	42
Obrázek 4.13	GUI, blok: Find glottal pulses.	43
Obrázek 4.14	GUI, blok: Show (compare) glottal flow.	44
Obrázek 4.15	GUI, blok: Export glottal flow data to.	45
Obrázek 4.16	Okno zobrazující: a) všechny nalezené pulzy, b) odstraněný chybný pulz číslo 5.	46
Obrázek 4.17	Okno zobrazující informace o zvoleném pulzu.	47
Obrázek 4.18	Okno zobrazující směrodatnou odchylku od průměrného pulzu.	47
Obrázek 4.19	a) porovnání výstupu metody IAIF a Sahoo na stejný řečový signál, b) porovnání výstupu metody IAIF a dat z EGG, c) porovnání výstupu metody Sahoo a dat z EGG.	48
Obrázek 4.20	Ukázka exportu dat do excelového souboru.	49
Obrázek 5.1	Histogram koeficientu špičatosti průměrného hlasivkového pulzu pro HC i PN.	51
Obrázek 5.2	Histogram koeficientu šikmosti průměrného hlasivkového pulzu pro HC i PN.	51
Obrázek B.1	Algoritmus pro určení fáze uzavření hlasivek z řečového signálu.	61

SEZNAM TABULEK

Tabulka 1.1	Řečová syntéza, publikace I.....	12
Tabulka 1.2	Řečová syntéza, publikace II.	13
Tabulka 1.3	Řečová syntéza, publikace III.	13
Tabulka 1.4	Rozpoznání mluvčího, publikace I.	13
Tabulka 1.5	Rozpoznání mluvčího, publikace II.	14
Tabulka 1.6	Rozpoznání mluvčího, publikace III.	14
Tabulka 1.7	Medicínské využití, publikace I.	14
Tabulka 1.8	Medicínské využití, publikace II.	15
Tabulka 1.9	Medicínské využití, publikace III.	15
Tabulka 5.2	Zjištěné hodnoty pro PN	50
Tabulka 5.1	Zjištěné hodnoty pro HC.....	50

ÚVOD

Každý známý kulturní jazyk na světě existuje ve dvou základních dorozumívacích tvarech, psané formě (písmo) a mluvené formě (řeč). Dává člověku možnost, díky které je schopen předání (popisu) svých pocitů, myšlenek, závěrů či jiných důležitých informací jinému člověku. Oba dva tvary si jsou rovnocenné, písemný tvar umožňuje „uložení“ informací a jejich následné přečtení kdykoliv v čase, například od osoby dalece vzdálené od nás. Mluvená forma je velice efektivní při komunikaci minimálně dvou osob nacházejících se blízko sebe, informace je předána hned při vyřčení a následně zaniká (pokud neuvažujeme doznívání akustického signálu ozvěnou, případně jeho nahrání).

Tato práce se zabývá mluvenou formou řeči a v ní zkoumá problematiku odhadu tzv. hlasivkového pulzu z řečového záznamu. Tento pulz je základním stavebním kamenem pro vytváření znělých zvuků, které vznikají pouze při vyslovování samohlásek. Analýzou těchto hlasivkových pulzů může být detekováno onemocnění hlasivek, odhadnutí psychického stavu člověka, dosažení přirozenějšího projevu syntetické řeči a jiných věcí, které jsou popsány dále v této práci.

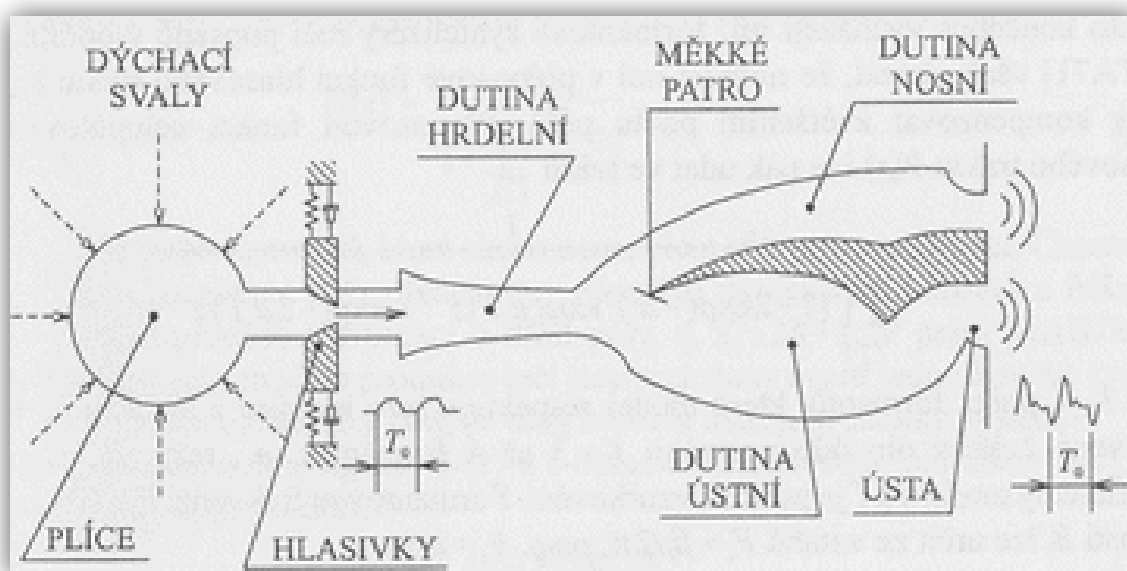
Jako krátké úseky řeči jsou v této práci myšleny signály, které obsahují jednu samohlásku (znělý úsek) trvající krátkou dobu (například 100 ms). Případně signály kde není mnoho znělých úseků, protože nastavování parametrů každého úseku pro výpočet by zabralo příliš času.

Jako dlouhé úseky řeči jsou v této práci myšleny signály s větším počtem znělých úseků. Při analýze dlouhých úseků se výpočetní metoda nastaví pouze na začátku a pracuje stále se stejnými parametry po celou dobu výpočtu, případně si je metoda sama vhodně upravuje.

1 VZNIK A VLASTNOSTI MLUVENÉ ŘEČI

1.1 Proces vytváření řeči člověkem

Lidské tělo nemá žádný orgán primárně zodpovědný za vytváření řečového (akustického) signálu, ovšem orgány zodpovědné za tvorbu řeči řadíme do skupiny řečových (artikulačních) orgánů, případně mluvidel (artikulátorů). Všechny řečové orgány spolu tvoří tzv. hlasivkový trakt, jehož přibližná délka od hrtanu až ke rtům je 17 cm a plocha příčného průřezu od 0 cm² (úplné uzavření traktu) do 20 cm² (úplné otevření traktu) [1]. Tvoří ho tři základní ústrojí dechové, hlasové a artikulační. Pro představu je níže znázorněn zjednodušený schématický model hlasového traktu (viz Obr 1.1).

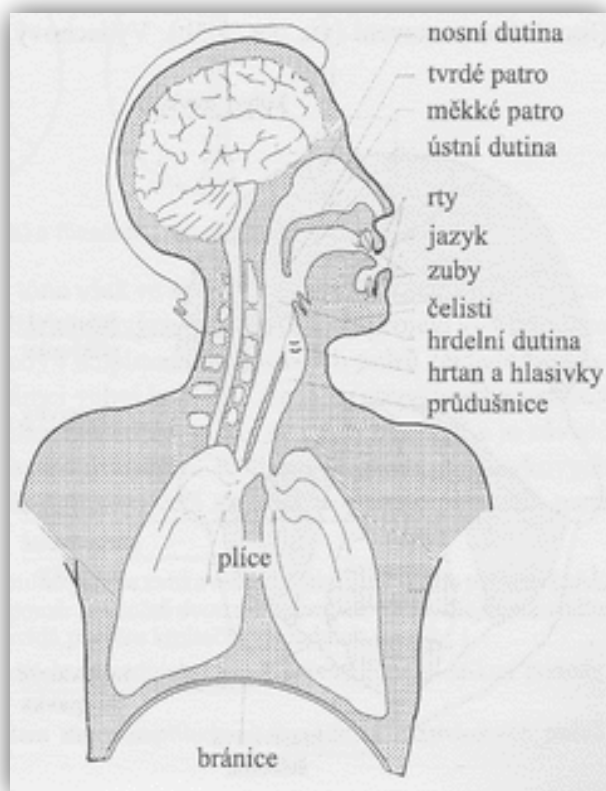


Obrázek 1.1 Schématický model hlasového ústrojí člověka [1].

1.1.1 Dechové ústrojí

Fundamentální zdroj energie pro řeč umístěný v hrudním koši. Tvořený přívodní dýchací cestou, plícemi a s nimi funkčně spjatými dýchacími svaly (bránicí) [1].

Vitální kapacita plic je hodnota udávající maximální rozdíl mezi úplně naplněnými a vypuštěnými plícemi, je asi 5 litrů. Přitom klidová kapacita plic dospělé osoby je přibližně 4 – 5 litrů a 1 – 2 litry tvoří zbytkovou kapacitu, která je vždy zachována. K produkci slyšitelné řeči je potřebné během několika sekund vytlačit z plic více než 0,5 litrů vzduchu. Při běžné řeči se spotřebuje polovina vitální kapacity plic, u velmi hlasité řeči až 80% kapacity. Nádechy doplňují vzduch (energii pro tvorbu řeči) do plic, v souvislé řeči se projevují jako pauzy [1].



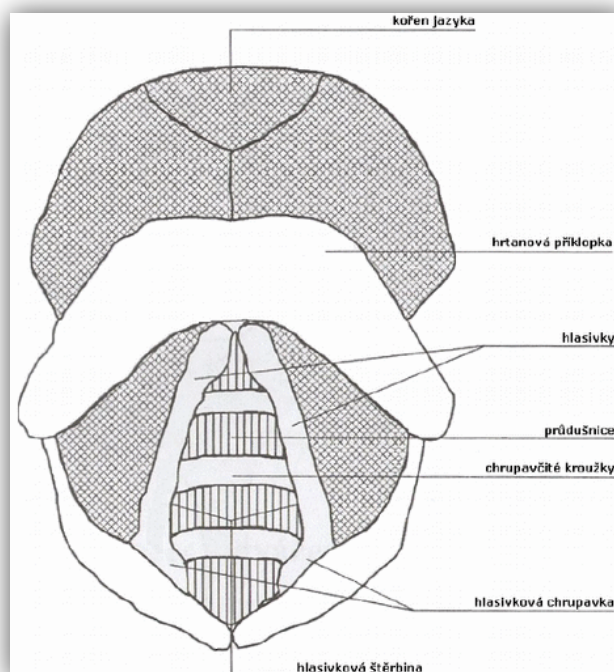
Obrázek 1.2 Hlasový trakt člověka [1].

1.1.2 Hlasové ústrojí

Část hlasového traktu kde dochází k samotnému vzniku hlasu. Nalézá se v hrtanu a jeho propojení s plícemi tvoří průdušnice.

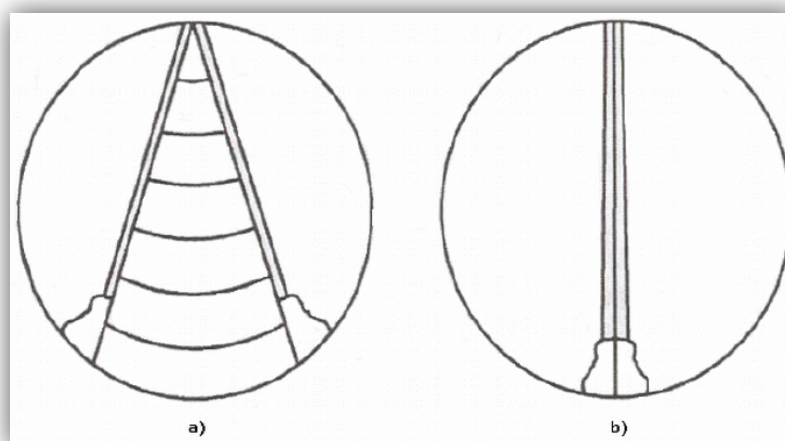
V hrtanové dutině (viz Obr 1.3) přímo za „ohryzkem“ v místě nejužšího průchodu se nalézají hlasivky, jsou složeny ze dvou slizničních ostrých řas. Prostor mezi hlasivkami tvoří tzv. hlasivkovou štěrbinu trojúhelníkového tvaru, ta pokud člověk mlčí je odkrytá, nalézá se v klidovém postavení (viz Obr 1.4 a)) a umožňuje průchod vzduchu potřebného k dýchání.

Při vytváření hlasu se hlasivková štěrbinu musí úplně uzavřít (stáhnout), nalézá se v tzv. hlasovém (fonačním) postavení (viz Obr 1.4 b)). Vydechovaný (vytlačovaný) proud vzduchu z plic postupuje bez jakéhokoli odporu průdušnicí do hrtanu, kde narazí na uzavřené (stažené) hlasivky. Ty se vlivem tlaku vzduchu stávají pružnými a začínají kmitat (jsou postupně otevírány a prudce uzavřeny).



Obrázek 1.3 Hrtan a hlasivky (pohled shora) [1].

Po průchodu hlasivkami se z původně homogenního proudu vzduchu stává proud s oblastmi kvant hustšího a řidšího vzduchu, vznikají zvukové vlny vnímané jako zvuk. Periodické opakování vzduchových pulsů je označováno termínem **základní (hlasivkový) tón**, frekvence kmitání hlasivek se nazývá **fundamentální frekvence** nebo **frekvence základního hlasivkového tónu** - označení F_0 .



Obrázek 1.4 Ukázka tvaru hlasivkové štěrbiny při klidu **a)** a fonačním postavením **b)** [1].

Při normální řeči je frekvence základního hlasivkového tónu u mužů v rozsahu 80 – 160 Hz s průměrem okolo 132 Hz, u žen 150 – 300 Hz s průměrem 223 Hz a u dětí 200 – 600 Hz [1]. Tato fundamentální frekvence není konstantní v delších časových

úsecích (vliv intonace), ani v kratších časových okamžicích, kde se mírné kolísání délky základní periody nazývá **jitter** a je závislí na duševním stavu mluvčího.

Znělé zvuky (samohlásky a znělé souhlásky) jsou tvořeny při fonačním postavením hlasivek, kmitání hlasivek je pravidelné a vzniklý zvuk má tónovou strukturu. **Neznělé zvuky** jsou tvořeny při klidovém postavení hlasivek, výsledný zvuk nemá žádný hlasivkový tón a je tvořen až pomocí artikulačního ústrojí.

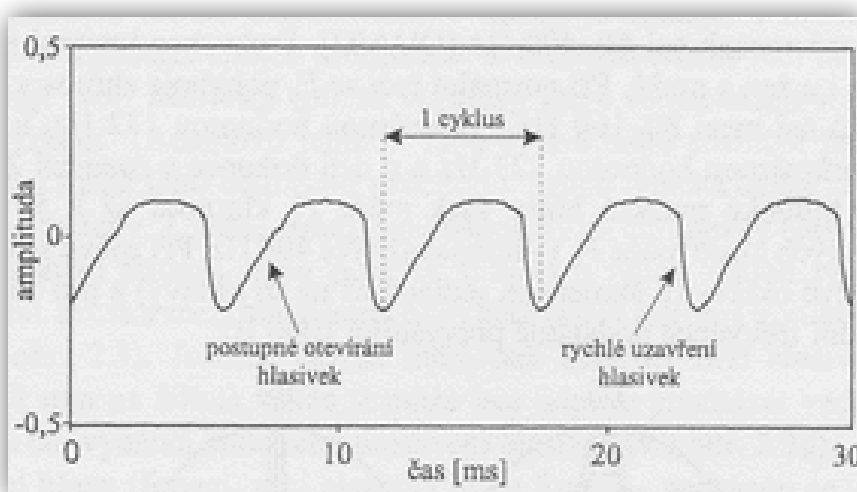
1.1.3 Artikulační ústrojí

Poslední část hlasového traktu, dochází zde k finální tvorbě zvuků charakterizující určitý kulturní mluvený jazyk.

Skládá se z **nadhrtanových dutin** (hrdelní, ústní a nosní) ty se na procesu tvorby řeči podílejí pasivně (nepohybují se). Naopak **artikulační orgány** jako jazyk, rty a měkké patro se na tvorbě řeči podílejí aktivním způsobem, svým pohybem ovlivňují velikost nadhrtanových dutin. Tento pohyb je během řeči plynulý, ovšem na velmi krátkou dobu zůstává soubor artikulátorů v jednom určitém postavení odpovídající dané určité řečové jednotce (samohlásce či souhlásce). Nejsložitějším a nejdůležitějším artikulačním orgánem, který nejvíce ovlivňuje rozložení nadhrtanových dutin je jazyk, orgán skládající se ze tří do jisté míry na sobě nezávislých částí hrot (špička), hřbet a kořen.

1.2 Přístrojové měření hlasivkových pulzů

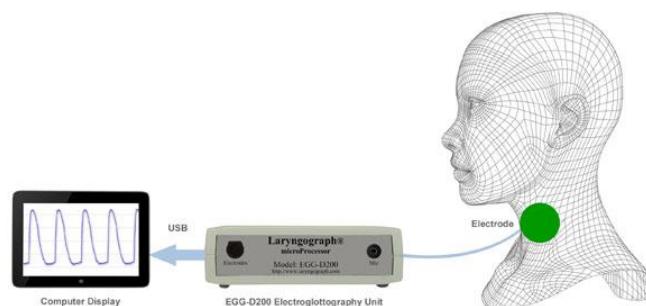
Pro měření činnosti hlasivek bylo vytvořeno vícero metod, některé z nich budou níže popsány. Ovšem nejčastěji se můžeme setkat s užitím měřicího přístroje elektroglossografu (EGG), někdy nazývaný jako laryngograf. Změřený průběh hlasivkového pulzu získaný pomocí více zmíněné metody můžeme vidět níže (viz Obr. 1.5).



Obrázek 1.5 Hlasivkový signál změřený elektroglossografem [1].

1.2.1 Elektrolottograf (EGG), laryngograf

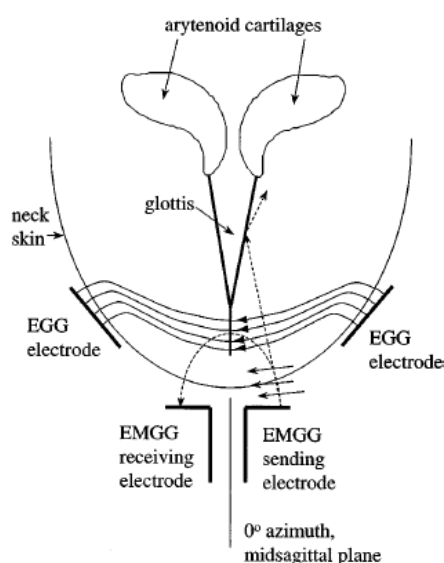
Je elektrické zařízení (viz Obr 1.6), pomocí něhož je možné provádět neinvazivní zjištění tvaru hlasivkového signálu. Pro měření se používají elektrody umístěné na povrchu krku v místě, kde se nachází hrtan. Mezi elektrodami prochází malý střídavý proud v řádu MHz. Jakmile se hlasivky rozkmitají, jsou neustále otevírány a zavírány přičemž se zároveň mění impedance mezi elektrodami. Ze změřené změny impedance je zjištěn hlasivkový signál.



Obrázek 1.6 Ukázka přístroje elektrolottografu s výstupními daty a umístěním elektrod na krku [2].

1.2.2 Elektromagnetický glottograf (EMGG)

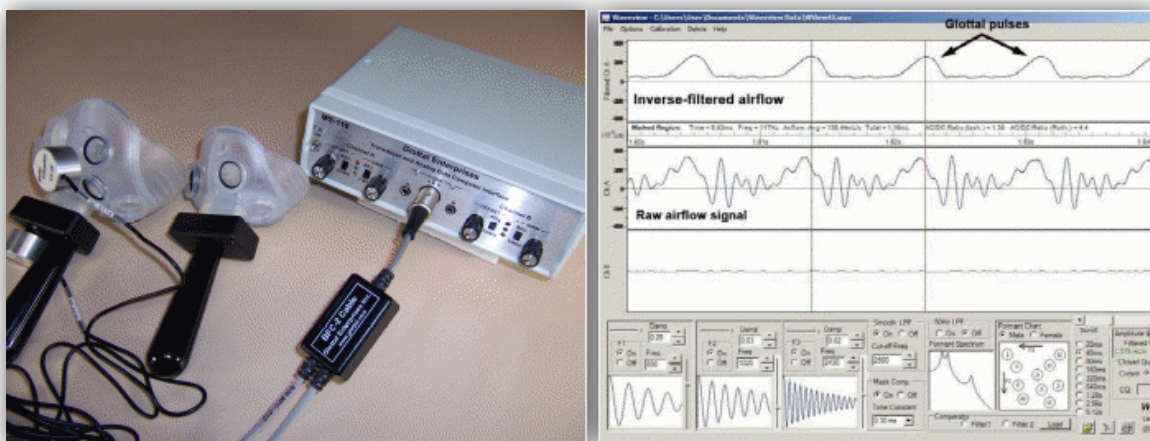
Funguje na stejném principu jako EGG, ovšem místo měřících elektrod umístěných na krku se pro detekci změny impedance použijí antény umístěné naproti hrtanu. Jedna anténa je vysílací a druhá přijímací, z Obrázku 1.7 znázorňující řez krkem v rovině hlasivek můžeme vidět, že vyzařovaná elektrická složka je rovinná. Pro porovnání můžeme vidět i umístění elektrod na krk pro měření EGG. Výhoda této metody je nulová potřeba dotyku s kůží subjektu pro měření. Nevýhodou je výrazné ovlivnění EMGG signálu způsobené pohybem stěny průdušnice [3].



Obrázek 1.7 Umístění antén EMGG a elektrod EGG na krku [3].

1.2.3 Měřicí sestava „Aeroview System“

Je kompaktní, neinvazivní měřicí sestava (viz Obr 1.8) vyvinutá firmou „Glottal Enterprises“ umožňující přesné zaznamenání proudění vzduchu v ústní i nosní dutině, subglottálního tlaku a nahrání řeči mikrofonem, tyto parametry jsou získány za pomoci OroNasální masky (OroNasal mask) [4]. Sestava obsahuje program „Waveview“ pro inverzní filtrování proudu vzduchu a zjištění hlasivkových pulzů.

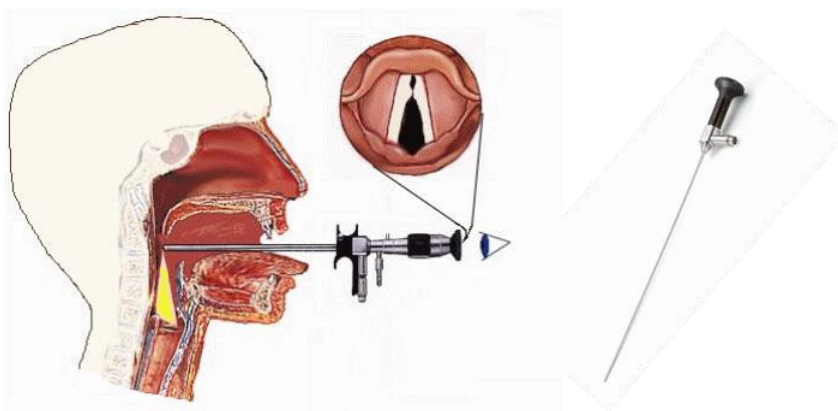


Obrázek 1.8 Vlevo měřicí sestava „Aeroview System“, vpravo program Waveview [4].

1.2.4 Videolaryngostroboskopie

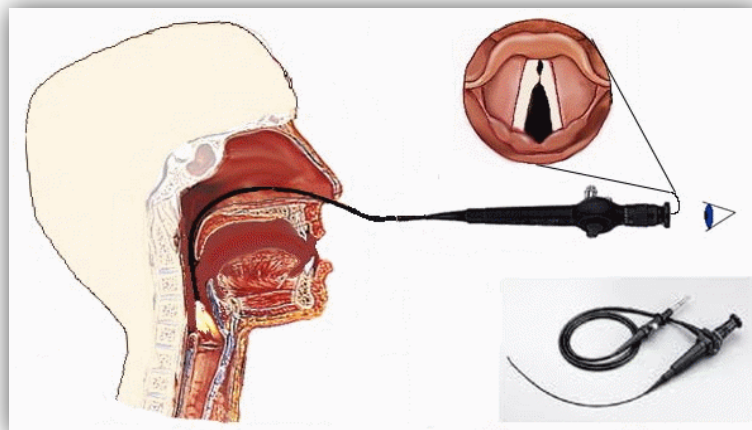
Metoda umožňuje kromě zjištění vibračních charakteristik hlasivek, také jejich přímé pozorování během tvorby zvuku za pomoci endoskopu. Používají se zde dva druhy (tvary) endoskopů [5].

První pevného (rigidního) tvaru (viz Obr 1.9), se vkládá do úst tak aby jeho konec byl až za zadní částí jazyka. Díky tomu má malá kamera umístěná na konci endoskopu přímý pohled na hlasivky.



Obrázek 1.9 Endoskop pevného (rigidního) tvaru [6].

Druhý ohebného (flexibilního) tvaru (viz Obr 1.10), umožňuje díky své ohebnosti a malé tloušťce zavedení skrze nosní dutinu. Na konci se opět nalézá malá kamera umožňující přímý pohled na hlasivky.

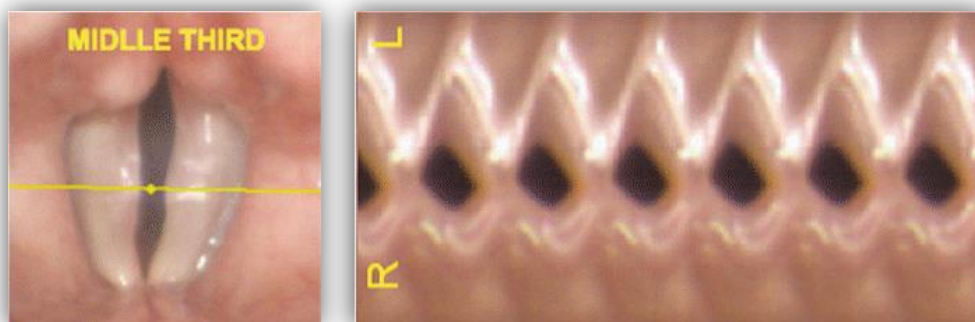


Obrázek 1.10 Endoskop ohebného (flexibilního) tvaru [6].

Protože chvění hlasivek probíhá s rychlostí frekvence základního hlasivkového tónu, nejsme schopni toto kmitání lidských okem ani normální kamerou postřehnout, využije se stroboskopického efektu [5]. Kdy se zdroj světla ozařující hlasivky v pravidelném intervalu přerušuje (rozsvěcuje a zhasíná) a vytvoří se plynulý obraz, ve kterém se hlasivky pohybují „pomalu“ rychlostí.

1.2.5 Videokymografie

Metoda vyvinutá českým (Jan G. Švec) a německým (Harm K. Schutte) vědcem v roce 1994. Speciálně upravená kamera na konci endoskopu snímá pouze velmi malou úzkou část hlasivek (jeden řádek) se snímkovací frekvencí 8000 Hz [7] (viz Obr 1.11). Díky tomu, mohou být zobrazeny (zjištěny) nepravidelné nebo krátce trvající děje, které při využití stroboskopického efektu nemohou být zaznamenány. Používá se pro doplnění stroboskopického nálezu.



Obrázek 1.11 Vlevo hlasivky s vyznačeným snímaným řádkem, vpravo výstup metody [8].

1.2.6 Vysokorychlostní videolaryngoskopie

Metoda kombinující vlastnosti videolaryngoskopie a videokimografie. Na konci endoskopu je umístěná kamera schopná snímání obrazu celých hlasivek s tak vysokou snímkovou frekvencí, aby bylo možné zobrazit i vysoce nepravidelné děje.

1.3 Akustické (programové) měření hlasivkových pulzů

Vedle přístrojového (většinou klinického) měření hlasivkových pulzů je možné zjistit jejich tvar i ze záznamu řečového signálu bez použití výše zmíněných měřicích přístrojů. Níže popsany software nevyžaduje žádné matematické ani programovací znalosti, k jeho použití dostačuje pouze řečový záznam zaznamenaný pomocí mikrofону, většinou ve formátu “.WAV“. Výhodou tohoto datového formátu je uložení nekomprimovaných dat z lineárně pulzně kódové modulace (LPCM), jedná o bezeztrátový záznam zvukového signálu.

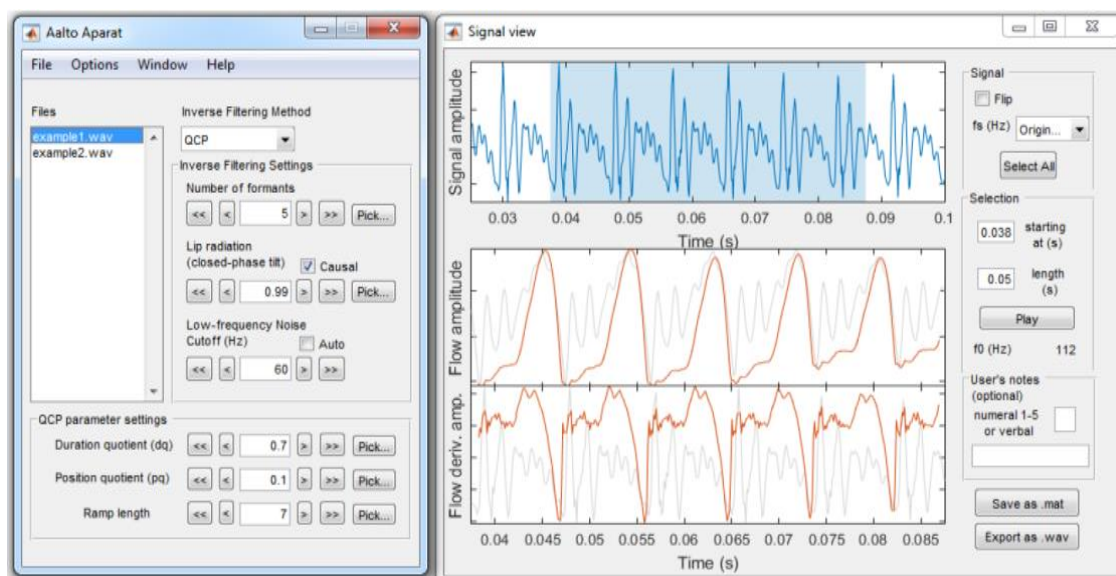
1.3.1 Aalto Aparat

Volně dostupný open-source program napsaný a pracující v interaktivním programovém prostředí MATLAB. Vyvinutý na Aalto-yliopistokse univerzitě elektrotechniky, katedře zpracování signálu a akustiky.

Pro inverzní filtrování hlasu jsou zde implementovány dvě metody schopné získání hlasivkových pulzů „quasi closed phase“ (QCP) analýza a „iterative adaptive inverse filtering“ (IAIF). Oba dva implementované algoritmy mají tři společné parametry pro nastavení inverzního filtrování následované specifickým nastavením vybrané metody. U QCP je možné ovlivnit výchozí hodnotu funkce pro váhovací křivku, pro IAIF se vybírá metoda pro modelování autoregresivního modelu.

Podle [9] poskytuje QCP lepší kvalitu inverzního filtrování v porovnání s IAIF, ovšem IAIF dokáže pracovat i signálem s malým poměrem SNR.

Dostupný na adrese: <http://research.spa.aalto.fi/projects/aparat/>



Obrázek 1.12 Ukázka prostředí programu Aalto Aparat [9].

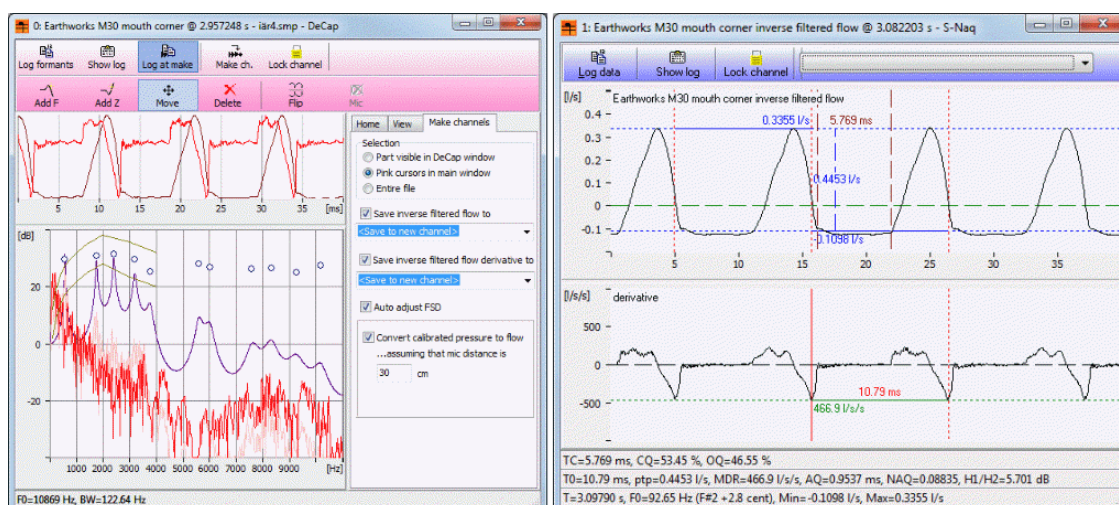
1.3.2 Sopran (DeCap)

DeCap je volně dostupný software pro inverzní filtrování zvukového signálu od společnosti **Tolvan Data**.

Tvůrce programu nikde neuvádí, jaká byla použita metoda inverzního filtrování, stejně tak nebylo možné nikde dohledat dokumentaci k programu. Bylo zjištěno, že pro správnou funkčnost musí být vstupní data ve formátu “.SMP“ a samostatný program DeCap je nyní začleněn do softwaru Sopran [10].

Sopran je stále se vyvíjející zvukový editor pro zpracování a analýzu zvukového záznamu od společnosti Tolvan Data. Stejně jako výše nebylo možné zjistit metodu pro inverzní filtrování a dokumentace k programu také nedostupná.

Dostupný na adrese: <http://www.tolvan.com/index.php?page=/sopran/sopran.php>



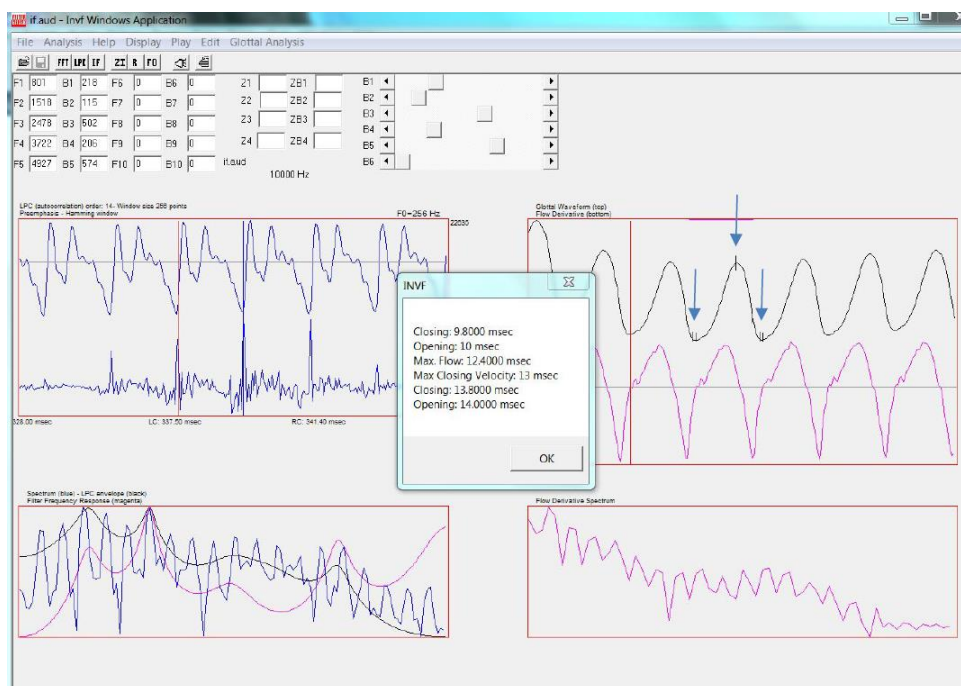
Obrázek 1.13 Ukázka inverzního filtrování v programu Sopran [10].

1.3.3 Inverse filtering (UCLA Bureau of Glottal Affairs)

Program „Inverse filtering“ je součástí balíčku obsahující software pro hlasovou syntézu a hlasovou analýzu. Všechny tyto programy byly vyvinuty na Kalifornské univerzitě v Los Angeles oddělením zabývajícím se záležitostmi okolo hlasivek, jsou licenčně dostupné jako shareware, zdrojové kódy jsou dostupné na vyžádání [11].

Pro výpočet hlasivkových pulzů se používá metoda popsaná v [12] (Javkin et al. (1987)).

Dostupný na adrese: <https://www.uclahealth.org/head-neck-surgery/bga/software>



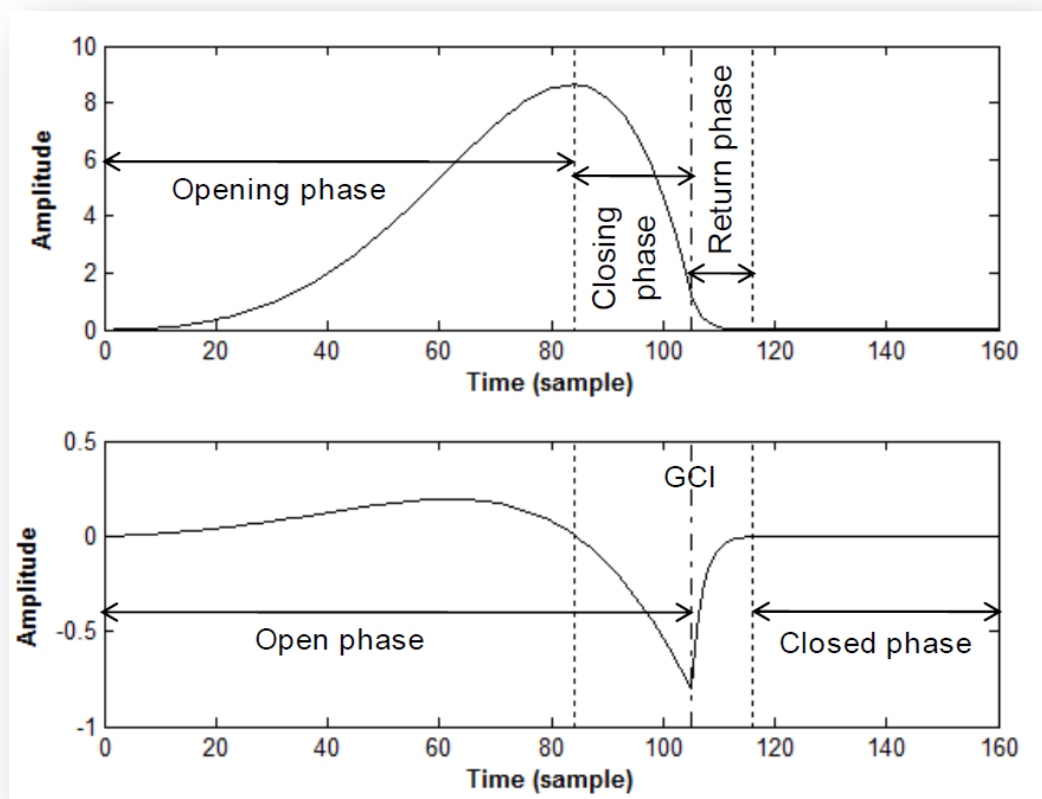
Obrázek 1.14 Ukázka inverzního filtrování v programu Inverse filtering (UCLA Bureau of Glottal Affairs) [11].

1.4 Využití hlasivkových pulzů

Tvar křivky hlasivkového pulzu udává rychlost proudění vzduchu skrze hlasivkovou štěrbinu. Pro popis těchto pulzů byly navrženy různé modely založené na analýze naměřených dat z fyziologických měření s odlišnou úrovní složitosti a jsou často pojmenovány po svých tvůrcích Rosenberg [13], Liljencrants-Fant (LF) [14], Fujisaki-Ljungqvist [15], Rosenberg++ [16].

Z nichž často používaný [17] je LF model ukázaný na Obrázku 1.15, kde jsou vyznačeny důležité fáze hlasivkového pulzu, fáze otevírání (opening phase) hlasivek, fáze uzavírání (closing phase) hlasivek a fáze návratu (return phase) hlasivek do jejich počátečního stavu. Pokud provedeme derivaci normálního (původního) průběhu LF modelu můžeme z grafu vyčíst místo kde dochází k uzavření hlasivek, toto místo přesně rozděluje pulz na část, kde jsou hlasivky otevřeny (open phase) a uzavřeny (close phase). Takto parametrizovaný hlasivkový pulzu nám dovoluje provést jeho následné

zpracování a analyzování pro další využití.



Obrázek 1.15 Hlasivkový pulz podle Liljencrants-Fant modelu [17].

V následujících podkapitolách budou zmíněny publikace zabývající se využitím parametrů hlasivkových pulzů.

1.4.1 Řečová syntéza

Tabulka 1.1 Řečová syntéza, publikace I.

Název:	Statistical parametric speech synthesis [18]
Autor:	Heiga Zen, Keiichi Tokuda, Alan W. Black
Místo, rok vydání:	Nagoya, Cambridge, Pittsburg, 2009
Obsah:	Práce se zabývá statickou parametrickou řečovou syntézou, konkrétně metodou skrytého Markovova modelu (hidden Markov model (HMM)), který je velmi efektivní pro syntézu přirozené řeči.
Odkaz:	http://linkinghub.elsevier.com/retrieve/pii/S0167639309000648

Tabulka 1.2 Řečová syntéza, publikace II.

Název:	Glottal Spectral Separation for Parametric Speech Synthesis [19]
Autor:	João P. Cabral, Steve Renals, Korin Richmond and Junichi Yamagishi
Místo, rok vydání:	University of Edinburgh, 2008
Obsah:	V této práci byla vyvinuta metoda hlasivkové spektrální separace (Glottal Spectral Separation (GSS)), která odstraňuje vliv působení hlasivkových pulzů na spektrální obálku řeči. Bylo zjištěno, že metodu GSS je možné použít pro změnu normálního řečového signálu na dýchavičný nebo chraptivý upravením parametrů LF modelu.
Odkaz:	https://www.era.lib.ed.ac.uk/handle/1842/3832

Tabulka 1.3 Řečová syntéza, publikace III.

Název:	Pitch transposition and breathiness modification using a glottal source model and its adapted vocal-tract filter [20]
Autor:	Degotex G., Roebel A., Rodet X.
Místo, rok vydání:	IEEE International Conference on Acoustics, Speech and Signal Processing, 2011
Obsah:	Článek se zabývá transpozicí hlasového tónu (například o jednu oktávu) a úpravou dýchavičného řečového signálu. S ohledem na to že si syntetizovaný hlas zachová větší přirozenosti při porovnání s obecnějšími metodami.
Odkaz:	http://ieeexplore.ieee.org/document/5947511/

1.4.2 Rozpoznání mluvčího

Tabulka 1.4 Rozpoznání mluvčího, publikace I.

Název:	An Overview of Automatic Speaker Recognition Technology [21]
Autor:	Reynolds Douglas A.
Místo, rok vydání:	IEEE International Conference on Acoustics Speech and Signal Processing, 2002
Obsah:	Publikace poskytuje základní přehled o této problematice. Její využití pro identifikaci (ze souboru známých hlasů) mluvčího nebo verifikaci (ověření identity) mluvčího, konkrétní aplikace (řízení přístupu, automatická personalizace,...) a základní techniky zpracování.
Odkaz:	http://ieeexplore.ieee.org/document/5745552/

Tabulka 1.5 Rozpoznání mluvčího, publikace II.

Název:	The Deterministic Plus Stochastic Model of the Residual Signal and Its Applications [22]
Autor:	Drugman Thomas, Dutoit Thierry
Místo, rok vydání:	IEEE Transactions on Audio, Speech, and Language Processing, 2012
Obsah:	Článek pojednává o Deterministickém plus Stochastickém Modelu (DSM) zbytkového signálu a jeho využití pro rozpoznávání mluvčího, protože jeho výsledky dosahují lepší míry rozpoznání než jiné metody založené na analýze hlasivek pulzů.
Odkaz:	http://ieeexplore.ieee.org/document/6029971/

Tabulka 1.6 Rozpoznání mluvčího, publikace III.

Název:	On separating glottal source and vocal tract information in telephony speaker verification [23]
Autor:	Kinnunen Tomi, Alku Paavo
Místo, rok vydání:	IEEE International Conference on Acoustics, Speech and Signal Processing, 2009
Obsah:	Tato práce se zaměřuje na rozpoznání mluvčího skrze telefonní linku, kde za pomoci metody inverzního hlasivkového filtrování zaměřuje na oddělení spektra hlasivkových pulzů a hlasového traktu.
Odkaz:	http://ieeexplore.ieee.org/document/4960641/

1.4.3 Medicínské využití

Tabulka 1.7 Medicínské využití, publikace I.

Název:	Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk [24]
Autor:	Ozdas A., Shiavi R.G., Silverman S.E., Silverman M.K., Wilkes D.M.
Místo, rok vydání:	IEEE Transactions on Biomedical Engineering, 2004
Obsah:	Článek se zabývá využitím řečových rysů jako jitteru a sklonu spektra hlasivkových pulzů jako indikátorů stavu člověka, u kterého je v blízké době zvýšené riziko sebevražedných sklonů.
Odkaz:	http://ieeexplore.ieee.org/document/1325813/

Tabulka 1.8 Medicínské využití, publikace II.

Název:	Vocal-Source Biomarkers for Depression: A Link to Psychomotor Activity [25]
Autor:	Quatieri Thomas F., Malyska Nicolas
Místo, rok vydání:	Lexington, 2012
Obsah:	Práce popisuje biologické značky a jejich extrakci z řečového signálu. Tyto značky mohou být použity pro monitorování závažnosti depresí, sledování účinků nové léčby nebo standardizování hodnocení depresí.
Odkaz:	https://www.ll.mit.edu/mission/cybersec/publications/publication-files/full_papers/2012_09_09_MalyskaN_Interspeech_FP.pdf

Tabulka 1.9 Medicínské využití, publikace III.

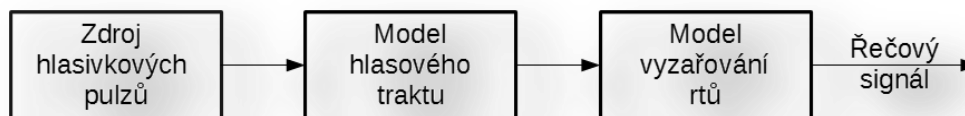
Název:	Glottal Flow Patterns Analyses for Parkinson's Disease Detection: Acoustic and Nonlinear Approaches [26]
Autor:	Bellalcázar-Bolaños E.A., Orozcoarroyave J. R., Vargas-Bonilla J. F., Haderlein T., Nöth E.
Místo, rok vydání:	Text, Speech and Dialogue, Proceedings, 2016
Obsah:	V článku je navržena metoda pro automatickou detekci Parkinsonovi nemoci z několika měření hlasivkových pulzů (jeho časově-frekvenčních vlastností) a nelineárního chování hlasivek.
Odkaz:	http://link.springer.com/10.1007/978-3-319-45510-5_46

2 ZÍSKÁNÍ HLASIVKOVÝCH PULZŮ ZE ZÁZNAMU ŘEČOVÉHO SIGNÁLU

Využívá se v případech, kdy je nemožné provést přístrojové měření pro získání tvaru hlasivkových pulzů z důvodu nedostupnosti řečníka, potřeby analyzování dříve pořízeného řečového záznamu či jiných důvodu nedovolujících přístrojové měření. Nebo pouhého faktu že pro zpracování signálu není potřeba přesná data získaná pomocí EGG a vystačí nám pouze „méně“ přesnější tvary pulzů získané za pomoci některých dále uvedených metod.

2.1 Metoda Iterativně adaptivního inverzního filtrování (Iterative Adaptive Inverse Filtering (IAIF))

Poloautomatická inverzně filtrovací metoda vyvinutá panem Paavo Alku a publikovaná v roce 1992 [27] využívá jako vstupní data řečový signál a generuje k nim odpovídající hlasivkové pulzy, blokový vývojový diagram použitého algoritmu IAIF je k dispozici na Obrázku 2.2. Metoda předpokládá následující model tvorby řeči (viz Obr. 2.1).



Obrázek 2.1 Model tvorby řečového signálu.

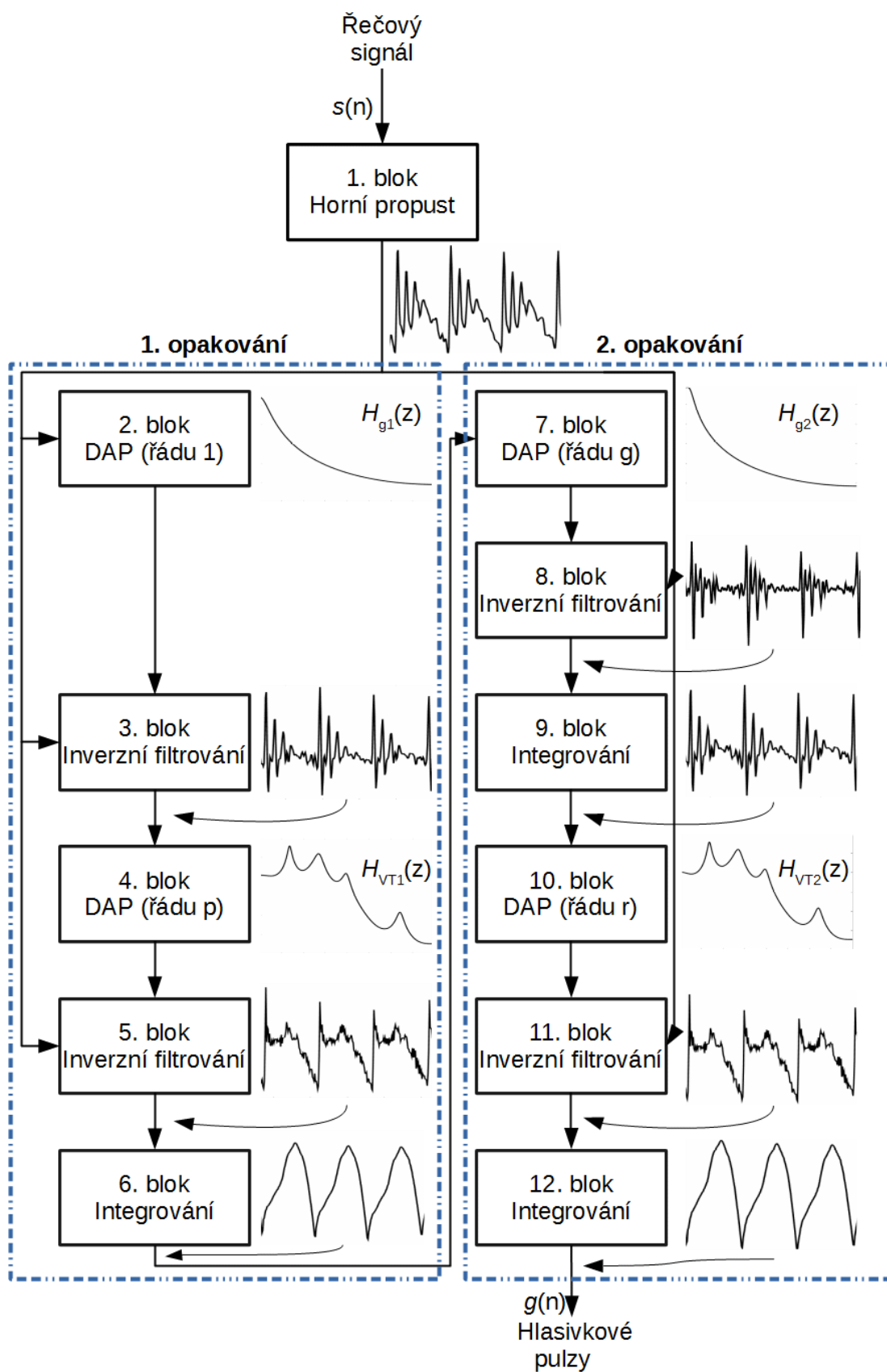
V roce 2005 se dočkala tato metoda vylepšení [28] spočívajícím v nahrazení bloků provádějících výpočet konvenčního lineárně prediktivního kódování (linear predictive coding (LPC)) za bloky provádějící výpočet metodou diskrétního all-pole modelování (discrete all-pole modeling (DAP)) [29]. Metoda DAP se zde používá pro snížení zkreslení způsobovaného harmonickou strukturou spektra řečového signálu při odhadu kmitočtů formantů.

Metoda pracuje ve dvou „opakováních“, proto má v názvu iterativní. V prvním opakování (fázi) se provede přibližný odhad vybuzení hlasivek, který je následně použit jako vstup do druhé fáze pro získání přesnějšího odhadu. Podrobnější popis metody po jednotlivých blocích je níže [28]:

1. **Horní propust** – vstupní signál je nejdříve filtrován horní propustí pro odstranění rušivých nízkofrekvenčních vlivů. Abychom filtrací nepřišli o relevantní informace, musí být mezní frekvence menší než fundamentální frekvence (F_0) řečového signálu. Pro filtrování byl vytvořen filtr typu FIR s mezní frekvencí menší než F_0 řečového signálu.
2. **DAP (řádu 1)** – jsou vypočítány parametry pro následující blok inverzního filtrování. V tomto bloku je proveden počáteční odhad společného účinku hlasivkového pulzu a efektu vyzařování rtů na řečové spektrum.

3. **Inverzní filtrování** – vstupní signál je filtrován filtrem s parametry z předchozího bloku. Díky tomu je odstraněn náklon řečového spektra způsobený spektrem budícího signálu a efektem vyzařování rtů.
4. **DAP (řádu p)** – výstupní data z předchozího bloku jsou použita pro výpočet DAP a získání modelu přenosové funkce hlasového traktu. Řád p DAP analýzy může být upraven uživatelem IAIF a odpovídá počtu formantů, které mají být modelovány v příslušném kmitočtovém pásmu.
5. **Inverzní filtrování** – vstupní signál je filtrován filtrem s parametry z předchozího bloku.
6. **Integrovaní** – na výstupních datech z předchozího bloku je provedena integrace pro zrušení efektu vyzařování rtů. V tomto bloku je proveden první odhad tvaru hlasivkových pulzů a dokončena první fáze (opakování).
7. **DAP (řádu g)** – výstupní data z předchozího bloku jsou použita pro výpočet DAP a získáme spektrální model vlivu hlasivkových pulzů na řečové spektrum. Řád g DAP analýzy je většinou volen mezi 2 – 4.
8. **Inverzní filtrování** – vstupní signál je filtrován filtrem s parametry z předchozího bloku. Díky tomu je odstraněn vliv příspěvků hlasivek na vstupní řečový signál.
9. **Integrovaní** – na výstupních datech z předchozího bloku je provedena integrace pro zrušení efektu vyzařování rtů.
10. **DAP (řádu r)** – výstupní data z předchozího bloku jsou použita pro výpočet DAP a získání nového modelu přenosové funkce hlasového traktu. Řád r DAP analýzy může být upraven uživatelem IAIF, běžně je ovšem roven hodnotě řádu p z bloku 4 ($r = p$).
11. **Inverzní filtrování** – vstupní signál je filtrován filtrem s parametry z předchozího bloku. Díky tomu je odstraněn vliv hlasového traktu na vstupní řečový signál.
12. **Integrovaní** – na výstupních datech z předchozího bloku je provedena integrace pro zrušení efektu vyzařování rtů a je získán konečný odhad hlasivkových pulzů. Konec metody IAIF.

Bloky provádějící integrování v této metodě jsou realizovány za použití integrátoru s netěsností (leaky integrator). Parametr netěsnosti ρ se může pohybovat v malém rozsahu pod 1, této práci byla jeho hodnota zvolena na 0,99.



Obrázek 2.2 Vývojový diagram algoritmu IAIF.

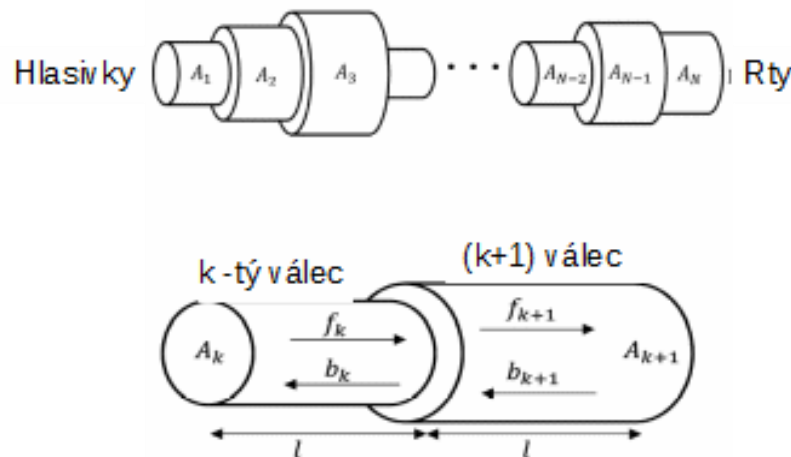
2.2 Sahoo -va metoda pro inverzní filtrování

Tato metoda je založena na rozloženém modelu hlasového traktu, kde je stavový model tvorby řečového signálu definován použitím válcového modelu hlasového traktu a LF-modelu hlasivkového pulzu. V tomto případě je tedy celý hlasový trakt modelován jako jediný systém se stavovými proměnnými rozloženými po celé délce válcového modelu. Kde vstupní signál tvoří hlasivkové pulzy a výstupní řečový signál je výstupem (odezvou) po projití hlasivkového pulzu celým systémem. Protože parametry vstupního signálu jsou neznámé a snažíme se je zjistit, stává se celý systém nelineárním.

Proces rozšířeného Kalmanova filtrování (extended Kalman filtering (EKF)) umožňuje linearizaci nelineárních systémových rovnic za použití Jacobianských matic pro lineární transformace [30].

2.2.1 Stavové rovnice popisující modelu hlasového traktu

Akustický válcový model hlasového traktu popisuje hlasový trakt jako spojení několika oddělených válcovitých trubek o stejné délce (viz Obr 2.3).



Obrázek 2.3 Válcový model hlasového traktu [30].

Proudění vzduchu skrze tyto válce se považuje za jednorozměrný (1-D) problém, který se dá vyjádřit za pomoci vlnovodových rovnic:

1. První válec navazující na hlasivky [30]:

$$f_1(t) = \frac{1}{2}(1 + r_g) u_g(t) + r_g b_1(t), \quad (2.1a)$$

$$b_1(t + \tau) = -r_1 f_1(t - \tau) + (1 - r_1) b_2(t), \quad (2.1b)$$

2. Pro válec $k = 2$ až $N-1$ [30]:

$$f_k(t) = (1 + r_{k-1}) f_{k-1}(t - \tau) + r_{k-1} b_k(t), \quad (2.2a)$$

$$b_k(t + \tau) = -r_k f_k(t - \tau) + (1 - r_k) b_{k+1}(t), \quad (2.2b)$$

3. Pro N -tý válec navazující na rty [30]:

$$f_N(t) = (1 + r_{N-1}) f_{N-1}(t - \tau) + r_{N-1} b_N(t), \quad (2.3a)$$

$$b_N(t + \tau) = -r_N f_N(t - \tau), \quad (2.3b)$$

4. Výstupní objem vzduchového signálu za rty [30]:

$$u_n(t) = (1 + r_N) f_N(t - \tau), \quad (2.4)$$

kde r_g, r_k, r_N jsou koeficienty odrazu popsány jako [30]:

$$r_g = \left(\frac{Z_g - \frac{\rho c}{A_1}}{Z_g + \frac{\rho c}{A_1}} \right), \quad (2.5)$$

$$r_k = \left(\frac{A_{k+1} - A_k}{A_k + A_{k+1}} \right), \quad (2.6)$$

$$r_N = \left(\frac{\frac{\rho c}{A_N} - Z_r}{\frac{\rho c}{A_N} + Z_r} \right), \quad (2.7)$$

kde A_k a A_{k+1} jsou plochy průřezu k -té a $(k+1)$ válce, $f_k(t)$ a $b_k(t)$ popisují proudění vzduchu vpřed a vzad skrze k -tý válec, ρ je hustota vzduchu, c je rychlost zvuku ve vzduchu, Z_g je impedance hlasivek, Z_r je impedance vyzařování rtů, u_g je vstupní signál z hlasivek do hlasového traktu. τ představuje čas potřebný k průchodu objemu vzduchu skrze jeden válec [30]:

$$\tau = \frac{l}{c}, \quad (2.9)$$

kde l je délka jedné válcovité sekce.

2.2.2 Lineární prostorový stavový model

Za pomoci vlnových 1-D rovnic byl vytvořen disktrétní stavový model pro tvorbu řečového signálu. Ve kterém byla jako stavová proměnná zvolena rychlost proudění objemu vzduchu skrze různé válce. Pro zachování stability výpočtu konečných diferenčních rovnic musí být splněna následující podmínka [30]:

$$c\Delta T \leq \Delta X, \quad (2.10)$$

kde ΔX je prostorový vzorkovací interval hlasového traktu a ΔT je vzorkovací interval řečového signálu. Dále předpokládejme, že doba vzorkovacího intervalu řečového signálu je rovna času potřebného pro průchod objemu vzduchu skrze válec ($\Delta T = \tau$).

Pokud v dříve definovaných 1-D vlnových rovnicích nahradíme t za n a τ nahradíme jedním časovým okamžikem, mohou být přepsány do tvaru [30]:

$$X[n+1] = A_n X[n] + B_n u_g[n], \quad (2.11a)$$

$$u_N[n+1] = C_n X[n]. \quad (2.11b)$$

Stavový vektor $X[n]$ a stavové matice A_n , B_n , C_n budou popsány níže.

$$X[n] = \begin{bmatrix} X_1[n] \\ X_2[n] \end{bmatrix}, \quad (2.12)$$

kde [30]:

$$X_1[n] = \begin{bmatrix} f_1[n-1] \\ f_2[n-1] \\ \vdots \\ f_N[n-1] \end{bmatrix}, \quad (2.12a)$$

$$X_2[n] = \begin{bmatrix} b_1[n] \\ b_2[n] \\ \vdots \\ b_N[n] \end{bmatrix}, \quad (2.12b)$$

kde $X_1[n]$ popisuje proudění vzduchu vpřed a $X_2[n]$ popisuje proudění vzduchu vzad (zpět) skrze válce 1 až N .

$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right], \quad (2.13)$$

kde [30]:

$$A_{11} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1+r_1 & 0 & \cdots & 0 & 0 \\ 0 & 1+r_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1+r_{N-1} & 0 \end{bmatrix}, \quad (2.13a)$$

$$A_{12} = \begin{bmatrix} r_g & 0 & \cdots & 0 & 0 \\ 0 & r_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & r_{N-2} & 0 \\ 0 & 0 & \cdots & 0 & r_{N-1} \end{bmatrix}, \quad (2.13b)$$

$$A_{21} = \begin{bmatrix} -r_1 & 0 & \cdots & 0 & 0 \\ 0 & -r_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -r_{N-1} & 0 \\ 0 & 0 & \cdots & 0 & -r_N \end{bmatrix}, \quad (2.13c)$$

$$A_{22} = \begin{bmatrix} 0 & 1 - r_1 & \cdots & \cdots & 0 \\ 0 & r_1 & 1 - r_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 - r_{N-1} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (2.13d)$$

kde A_{11} vztahuje rychlosti proudění vzduchu směrem vpřed v čase n až $n-1$, A_{12} vztahuje rychlosti proudění vzduchu směrem vpřed v čase n a rychlosti proudění vzduchu směrem vzad (zpět) v čase n , A_{21} vztahuje rychlosti proudění vzduchu směrem vzad (zpět) v čase $n+1$ a rychlosti proudění vzduchu směrem vpřed v čase $n-1$, A_{22} vztahuje rychlosti proudění vzduchu směrem vzad (zpět) v čase n až $n-1$.

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad (2.14)$$

kde [30]:

$$B_1 = \begin{bmatrix} \frac{1}{2}(1 + r_g) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (2.14a)$$

$$B_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (2.14b)$$

kde B_1 se vztahuje k objemu rychlosti vpřed se vstupním hlasivkovým signálem, B_2 je provázání zpětného proudění se vstupním signálem.

$$C = [C_1 \mid C_2], \quad (2.15)$$

$$C_1 = [0 \quad 0 \quad \cdots \quad 0 \quad 1 + r_N], \quad (2.15a)$$

$$C_2 = [0 \quad 0 \quad 0 \quad \cdots \quad 0], \quad (2.15a)$$

kde C_1 a C_2 spojuje výstupní řečový signál s vlnami putujícími směrem vpřed a vzad (zpět).

Takovýto zápis stavových vlnových rovnic umožňuje současně provést aktualizaci všech stavových proměnných při příchodu každého nového vzorku $u_g[n]$. Všechny stavové proměnné ve výše zmíněném modelu jsou ve formě objemové rychlosti vzduchu, ale řečový signál je zaznamenán ve formě tlakové změny signálu. Převod mezi těmito dvěma formami je u člověka řešen efektem vyzařování rtů, ten se modeluje pomocí derivace řečového signálu [30].

Protože se jako vstupní data do systému zmíněného v rovnicích (2.11) používá derivovaný hlasivkový pulz, jehož tvar je při výpočtu neznámý, je pro získání stavových rovnic použit LF-model derivovaného hlasivkového pulzu [30].

2.2.3 Stavové rovnice z modelu hlasivkového pulzu

Matematický výraz pro popis derivovaného hlasivkového signálu podle LF-modelu (viz Obr 2.4) je následující [30]:

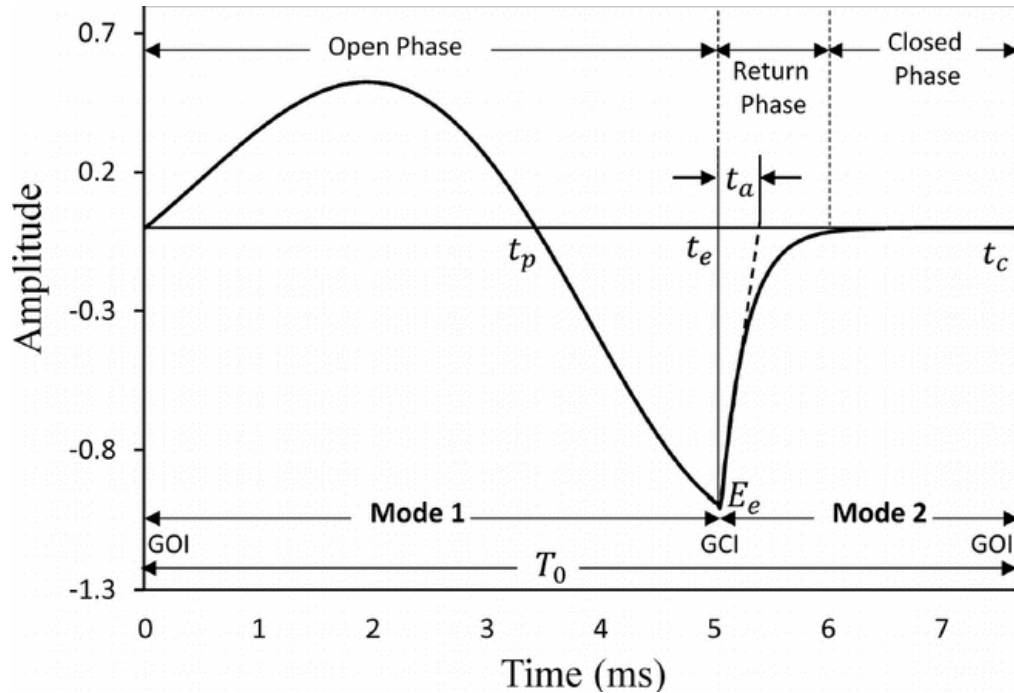
$$E(t) = E_0 e^{\alpha t} \sin(\omega_g t) \quad ; \quad t < t_e, \quad (2.16a)$$

$$E(t) = -\frac{E_e}{\varepsilon t_a} [e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)}] \quad ; \quad t_e < t < t_c, \quad (2.16b)$$

s omezením [24]:

$$\int_0^{T_0} E(t) dt = 0, \text{ pro } T_0 = t_c = \frac{1}{F_0}, \quad (2.17a)$$

$$\varepsilon t_a = 1 - e^{-\varepsilon(t_c-t_e)}. \quad (2.17b)$$



Obrázek 2.4 Derivovaný hlasivkový pulz LF-modelu s označením důležitých okamžiků [24].

Kde t_e je okamžik uzavření hlasivek ke kterému dochází při negativní špičce E_e , t_p je časový okamžik který odpovídá špičce v hlasivkovém pulzu a derivace hlasivkového pulzu je v nule, t_a souvisí s dobou trvání fáze návratu, t_c je časový okamžik při kterém se hlasivky začínají znovu otevírat (po skončení fáze uzavření). T_0 a F_0 je označení základní periody a frekvence. E_0 , ω_g , a α popisují tvar signálu během fáze otevření, E_e a ε popisují tvar signálu během fáze návratu. ε , α a E_0 jsou určeny z rovnic omezení (2.17).

Je-li možné derivovaný hlasivkový signál částečně rozlišovat rozdělujeme se do dvou separátních časových úseků (režimů): Režim 1 ($t < t_e$) a Režim 2 ($t_e < t < t_c$). Diskrétní stavové rovnice modelu hlasivkových pulzů pro oba režimy jsou získány následovně [30]:

Režim 1 ($t < t_e$): nechť $u_1(t) = E_0 e^{\alpha t} \sin(\omega_g t)$ a $u_2(t) = u_1(t)$

$$\dot{u}_1(t) = \alpha u_1(t) + \omega_g E_0 e^{\alpha t} \cos(\omega_g t), \quad (2.18a)$$

$$\dot{u}_2(t) = 2\alpha u_2(t) - (\omega_g^2 + \alpha_g^2) u_1(t), \quad (2.18b)$$

Nyní se použije dopředný rozdílový vzorec pro derivování ($\dot{u}(t) = \frac{u(t+\tau) - u(t)}{\tau}$) na rovnice (2.18) a získáme výrazy pro $u_1(t + \tau)$ a $u_2(t + \tau)$ [30]:

$$u_1(t + \tau) = u_1(t) + \tau u_2(t), \quad (2.19a)$$

$$u_2(t + \tau) = -(\omega_g^2 + \alpha_g^2) \tau u_1(t) + (1 + 2\alpha\tau) u_2(t). \quad (2.19b)$$

Režim 2 ($t_e < t < t_c$): nechť $u_1(t) = -\frac{E_e}{\varepsilon t_a} [e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)}]$ a $u_2(t) = u_1(t)$

$$\dot{u}_1(t) = -\varepsilon u_1(t) + \frac{E_e}{t_a} e^{-\varepsilon(t_c-t_e)}, \quad (2.20a)$$

$$\dot{u}_2(t) = -\varepsilon u_1(t). \quad (2.20b)$$

Nyní se použije stejný postup jako v Režimu 1 na rovnice (2.20) a získáme výrazy pro $u_1(t + \tau)$ a $u_2(t + \tau)$ [30]:

$$u_1(t + \tau) = u_1(t) + \tau u_2(t), \quad (2.21a)$$

$$u_2(t + \tau) = (1 - \tau\varepsilon) u_2(t), \quad (2.21b)$$

2.2.4 Nelineární prostorový stavový model

Po vložení stavových rovnic hlasivkového signálu do lineárního prostorového stavového modelu (2.11), může být výsledný stavový prostorový model zapsán v nelineární formě [30]:

$$\underline{x}_{n+1} = A_n(\underline{\theta}_n)\underline{x}_n + \underline{\xi}_n, \quad (2.22a)$$

$$u_n = C_n(\underline{\theta}_n)\underline{x}_n + \eta_n, \quad (2.22b)$$

kde \underline{x}_n je stavový vektor, $\underline{\theta}_n$ je vektor parametrů a u_n je zaznamenaný řečový signál. $\underline{\xi}$ a η reprezentují bílé Gaussovske šumové sekvence sdružené s tímto systémem. Veličiny $A_n(\underline{\theta}_n)\underline{x}_n$ a $\underline{\theta}_n$ budou pro každý režim jiné z důvodu rozdílných stavových rovnic použitých pro Režim 1 a 2 [30]:

Režim 1 ($t < t_e$):

$$\underline{x}_n = \begin{bmatrix} f_1[n-1] \\ f_2[n-1] \\ \vdots \\ f_N[n-1] \\ \hline b_1[n] \\ b_2[n] \\ \vdots \\ b_N[n] \\ \hline u_1[n] \\ u_2[n] \end{bmatrix}, \quad \underline{\theta}_n = \begin{bmatrix} r_g[n] \\ r_1[n] \\ r_2[n] \\ \vdots \\ r_{N-1}[n] \\ r_N[n] \\ \omega_g[n] \\ \alpha[n] \end{bmatrix}, \quad (2.23)$$

$$A_n(\underline{\theta}_n)\underline{x}_n = \begin{bmatrix} r_g b_1[n] + \frac{1}{2}(1 + r_g)u_1[n] \\ (1 + r_1)f_1[n-1] + r_1 b_2[n] \\ \vdots \\ (1 + r_{N-1})f_{N-1}[n-1] + r_{N-1}b_N[n] \\ \hline -r_1 f_1[n-1] + (1 - r_1)b_2[n] \\ -r_2 f_2[n-1] + (1 - r_2)b_3[n] \\ \vdots \\ -r_{N-1}f_{N-1}[n] + (1 - r_{N-1})b_N[n] \\ -r_N f_N[n-1] \\ \hline u_1[n] + \tau u_2[n] \\ \hline -(\omega_g^2 + \alpha_g^2)\tau u_1[n] + (1 + 2\alpha\tau)u_2[n] \end{bmatrix}, \quad (2.24)$$

Režim 2 ($t_e < t < t_c$):

Všechny elementy stavových matic kromě těch, které odpovídají výrazům $u_g[n]$ zůstávají stejné. Dva poslední elementy z matice $A_n(\underline{\theta}_n)\underline{x}_n$ jsou nahrazeny diskrétními tvary rovnic (2.21a) a (2.21b). Stejně tak i dva poslední elementy matice $\underline{\theta}_n$ jsou nahrazeny ϵ .

Protože koeficienty odrazu modelu hlasového traktu a parametry hlasivkového signálu jsou neznáme, je pro odhad těchto parametrů a stavů použito EKF. Ta se aplikuje zvlášť na každý režim stavového prostorového modelu.

2.2.5 Realizace rozšířeného Kalmanova filtrování (EKF)

Pro odhad parametrů stavů ve stavovém prostorovém modelu (2.22), je stavový vektor rozšířen o neznámé parametry [30]:

$$\begin{bmatrix} \underline{x}_{n+1} \\ \underline{\theta}_{n+1} \end{bmatrix} = \begin{bmatrix} A_n(\underline{\theta}_n) \underline{x}_n \\ \underline{\theta}_n \end{bmatrix} + \begin{bmatrix} \underline{\xi}_n \\ \underline{\zeta}_n \end{bmatrix}, \quad (2.25a)$$

$$u_n = [C_n(\underline{\theta}_n) \ 0] \begin{bmatrix} \underline{x}_n \\ \underline{\theta}_n \end{bmatrix} + \eta_n, \quad (2.25b)$$

kde $\underline{\zeta}$ reprezentuje Gaussovske sekvence bílého šumu spojené s parametry.

1. Inicializace stavů a parametrů pro \underline{x}_0 , $\underline{\theta}_0$. Také inicializace chyby kovarianční matice P na $P_{0,0}$.
2. Pro každý nový vzorek n řečového signálu vypočítat [30]:

$$\begin{bmatrix} \underline{x}_{n|n-1} \\ \underline{\theta}_{n|n-1} \end{bmatrix} = \begin{bmatrix} A_{n-1}(\underline{\theta}_{n-1}) \underline{x}_{n-1} \\ \underline{\theta}_{n-1} \end{bmatrix}, \quad (2.26)$$

$$\begin{aligned} P_{n,n-1} &= \begin{bmatrix} A_{n-1}(\underline{\theta}_{n-1}) & \frac{\partial}{\partial \underline{\theta}} [A_{n-1}(\underline{\theta}_{n-1}) \underline{x}_{n-1}] \\ 0 & I \end{bmatrix} P_{n-1,n-1} \cdot \\ &\cdot \begin{bmatrix} A_{n-1}(\underline{\theta}_{n-1}) & \frac{\partial}{\partial \underline{\theta}} [A_{n-1}(\underline{\theta}_{n-1}) \underline{x}_{n-1}] \\ 0 & I \end{bmatrix}^T + \\ &+ \begin{bmatrix} Q_{n-1} & 0 \\ 0 & S_{n-1} \end{bmatrix}, \end{aligned} \quad (2.27)$$

kde $Q_n = \text{Var}(\underline{\xi}_n)$ a $S_n = \text{Var}(\underline{\zeta}_n)$, následně [30]:

$$\begin{aligned} G_n &= P_{n,n-1} [C_n(\underline{\theta}_{n|n-1}) \ 0]^T \cdot \\ &\cdot \left[[C_n(\underline{\theta}_{n|n-1}) \ 0] P_{n,n-1} [C_n(\underline{\theta}_{n|n-1}) \ 0]^T + R_n \right]^{-1} \end{aligned} \quad (2.28)$$

3. Aktualizace dat stavové matice, matice parametrů a chyby kovarianční matice [24]:

$$P_{n,n} = [I - G_n [C_n(\underline{\theta}_{n|n-1}) \ 0]] P_{n|n-1}, \quad (2.29)$$

$$\begin{bmatrix} \underline{x}_n \\ \underline{\theta}_n \end{bmatrix} = \begin{bmatrix} \underline{x}_{n|n-1} \\ \underline{\theta}_{n|n-1} \end{bmatrix} + G_n(u_n - C_n(\underline{\theta}_{n|n-1})\underline{x}_{n|n-1}), \quad (2.30)$$

kroky 2 až 3 se opakují pro všechna n ($n = 1, 2, 3, \dots$). Pro Režim 1 modelu je sestavení Jacobiánské matice parametru $\frac{\partial}{\partial \underline{\theta}} [A_n(\underline{\theta}_n)\underline{x}_n]$ z rovnice (2.27) následující [30]:

$$\begin{bmatrix} b_1 + \frac{1}{2}u_g[n] & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & f_1 + b_2[n] & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & f_2 + b_3[n] & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & f_{N-1} + b_N[n] & 0 & 0 \\ \hline 0 & -(f_1 + b_2[n]) & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -(f_2 + b_3[n]) & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(f_{N-1} + b_N[n]) & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -f_N & 0 \\ \hline 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & xy \end{bmatrix}, \quad (2.31a)$$

kde $x = -2\omega_g\tau u_1[n]$ a $y = -2\alpha\tau u_1[n] + 2\tau u_2[n]$, pro Režim 2 modelu je sestavení Jacobiánské matice následující [30]:

$$\begin{bmatrix} b_1 + \frac{1}{2}u_g[n] & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & f_1 + b_2[n] & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & f_2 + b_3[n] & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & f_{N-1} + b_N[n] & 0 & 0 \\ \hline 0 & -(f_1 + b_2[n]) & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -(f_2 + b_3[n]) & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(f_{N-1} + b_N[n]) & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -f_N & 0 \\ \hline 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & -\tau u_2[n] \end{bmatrix}. \quad (2.31b)$$

2.2.6 Metoda odhadu hlasivkového pulzu

Prvním důležitým bodem této metody je zjištění pozice t_e při nichž dochází k uzavření hlasivek (GCI) a pozice t_c otevření hlasivek (GOI). Tyto pozice vytyčují hranice pro použití Režimu 1 a 2.

Následuje inicializace EKF, počáteční hodnoty stavového vektoru \underline{x}_0 byly nastaveny na nulu (není žádné proudění vzduchu), koeficienty odrazu v hlasovém traktu (r_1 až r_{N-1}) vektoru $\underline{\theta}_0$ byly nastaveny na hodnoty získané z rovnice (2.26) vycházející z naměřených dat [31] a měly by odpovídat korespondujícím samohláskám. r_g a r_N mají kladnou menší hodnotu blízkou, ale nerovnou jedné. ω_g , α a ϵ byla přiřazena náhodná kladná hodnota menší než jedna [30]. Diagonální matice Q hodnoty 0,001 má rozměr délky vektoru \underline{x}_0 , diagonální matice S hodnoty 0,01 má délku vektoru $\underline{\theta}_0$, jednotková matice P má rozměr délky vektorů \underline{x}_0 a $\underline{\theta}_0$ ($\underline{x}_0 + \underline{\theta}_0$), parametr R má hodnotu 0,01.

Během průběhu celé metody musí být splněna podmínka $c\Delta T \leq \Delta X$ (2.10). (Ukázka splnění podmínky samohlásky /a/: válcový model se skládá ze 44 sekcí s průměrnou délkou celého hlasového traktu 17,46 cm, odpovídající velikost prostorového vzorkování ΔX je 0,397 cm a za předpokladu rychlosti vzduchu 33 400 cm/s by měla být vzorkovací frekvence ($1/\Delta T$) pro splnění podmínky minimálně 84 170 Hz.) Její splnění je zajištěno nahráním signálu se vzorkovacím kmitočtem 88 200 Hz nebo převzorkováním řečového signálu na tento kmitočet.

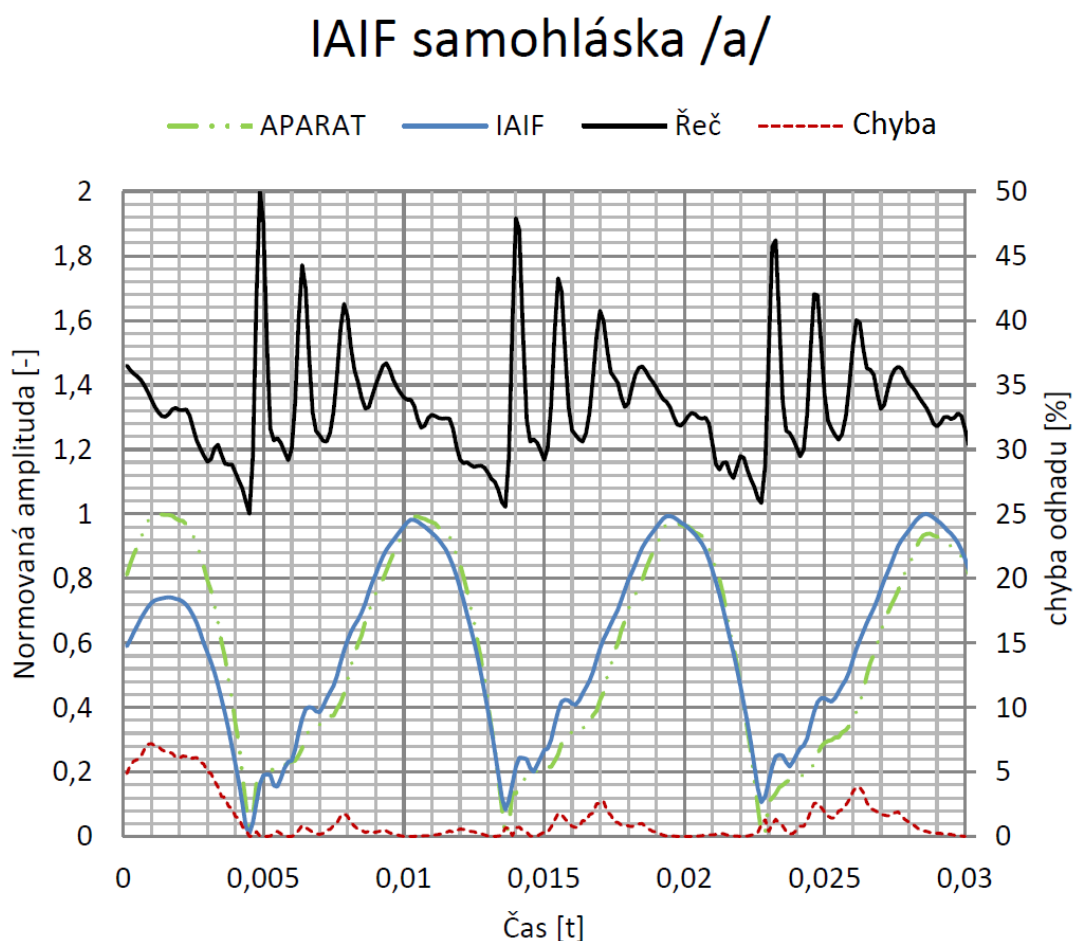
3 ODHAD TVARŮ HLASIVKOVÝCH PULZŮ

V této kapitole bude pomocí výše zmíněných metod, které byly naprogramovány ve vývojovém prostředí MATLAB proveden odhad tvaru hlasivkových pulzů z řečového signálu sejmutého za pomoci mikrofону.

3.1 IAIF metoda

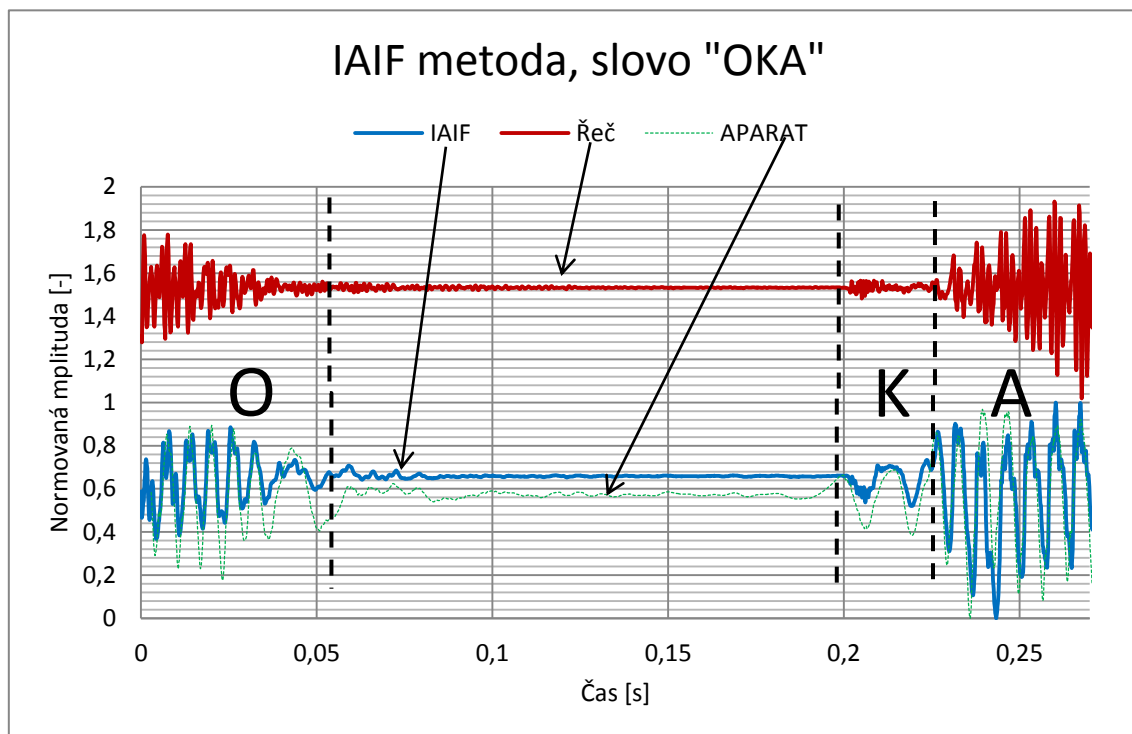
Metoda iteračně adaptivního inverzního filtrování (IAIF) používá odhad parametrů hlasového traktu pro inverzní filtraci řečového signálu, je často využívána pro odhad tvaru hlasivkových pulzů z důvodu její malé časové náročnosti pro výpočet. Průměrná doba výpočtu záznamu o délce 100 ms je 0,30 s.

Funkčnost metody byla ověřena pomocí programu Aparat u něhož byl i krátký vzorový záznam samohlásky /a/ ve formátu “.WAV“. Zjištěné hlasivkové pulzy pomocí této metody jsou na Obrázku 3.1.



Obrázek 3.1 Odhadnuté hlasivkové pulzy samohlásky /a/, metoda IAIF. Pro názornost byl na řečový signál aplikován offset.

Po ověření funkčnosti této metody na krátkém neměnném úseku, byla metoda podrobena zkoušce odhadu hlasivkového pulzu ze záznamu vlastního hlasu při vyřčení slova „OKA“. Odhadnuté pulzy jsou na Obrázku 3.2.



Obrázek 3.2 Odhadnuté hlasivkové pulzy slova „OKA“, metoda IAIF. Pro názornost byl na řečový signál aplikován offset.

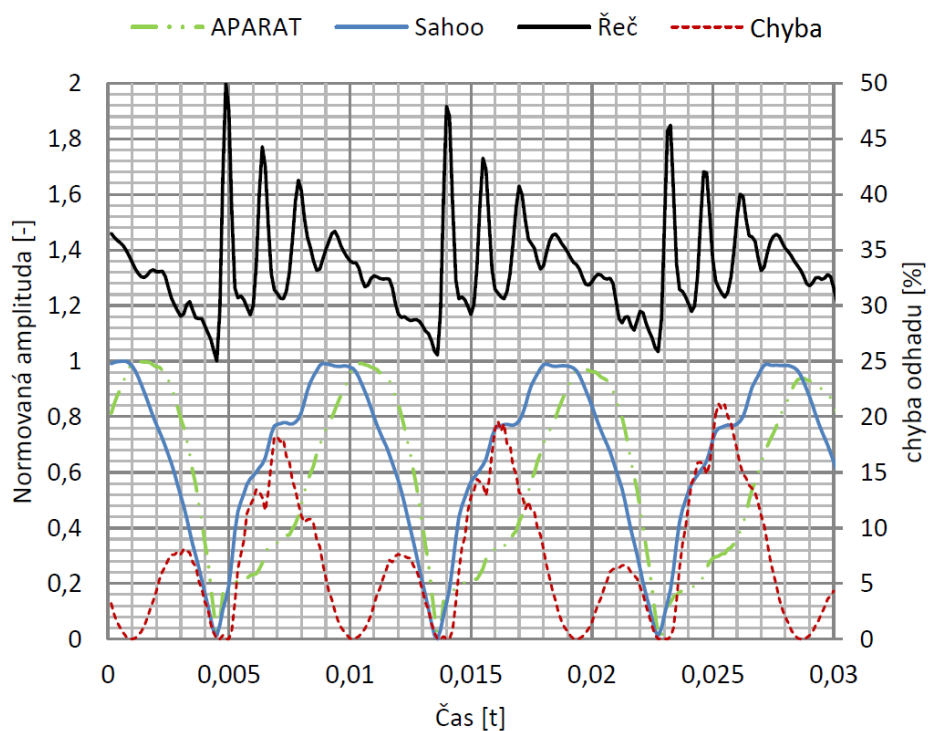
3.2 Sahoo –va metoda

Metoda navržená panem Sahoo používá LF model hlasového traktu, který je ovlivňován parametry řečového signálu pro dosažení nejdůvěryhodnějšího popisu hlasivkového pulzu. Průměrná doba výpočtu záznamu o délce 100 ms je 30 s.

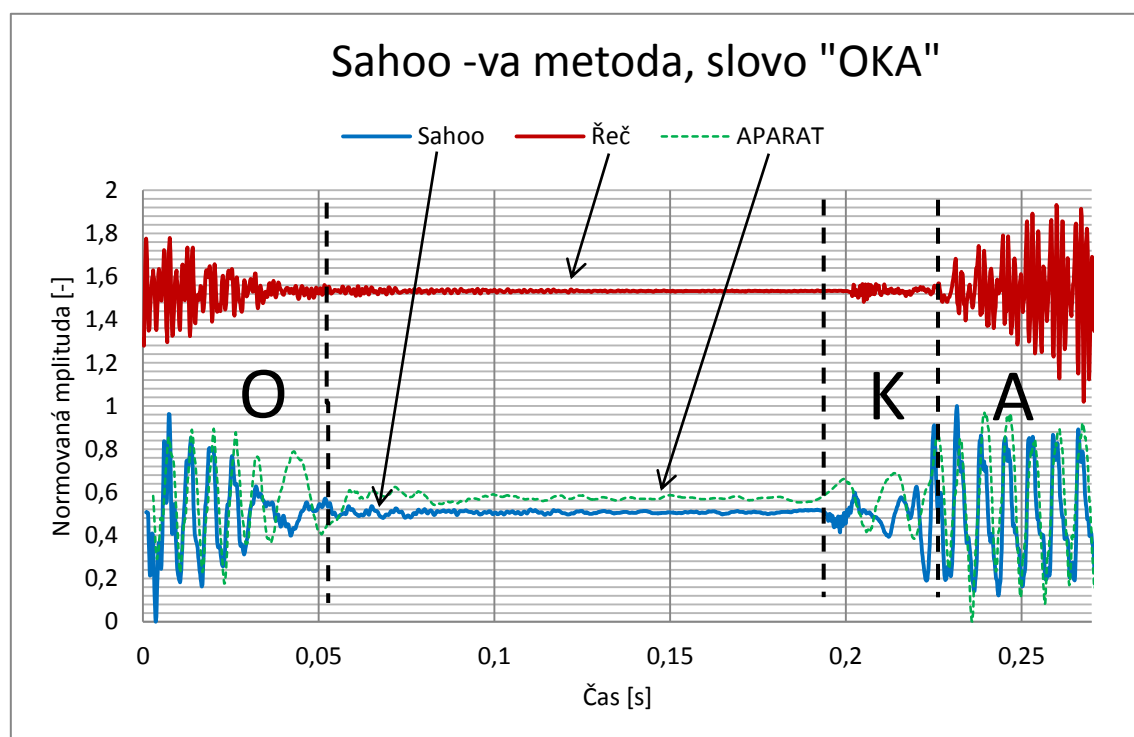
Funkčnost metody byla znovu ověřena pomocí programu Aparat a krátkého vzorového záznamu samohlásky /a/ ve formátu „.WAV“. Zjištěné hlasivkové pulzy pomocí této metody jsou na Obrázku 3.3.

Po ověření funkčnosti této metody na krátkém neměnném úseku, byla metoda podrobena zkoušce odhadu hlasivkového pulzu ze záznamu vlastního hlasu při vyřčení slova „OKA“. Odhadnuté pulzy jsou na Obrázku 3.4.

Sahoo samohláska /a/



Obrázek 3.3 Odhadnuté hlasivkové pulzy samohlásky /a/, Sahoo –va metoda. Pro názornost byl na řečový signál aplikován offset.



Obrázek 3.4 Odhadnuté hlasivkové pulzy slova „OKA“, Sahoo –va metoda. Pro názornost byl na řečový signál aplikován offset.

4 POPIS GRAFICKÉHO UŽIVATELSKÉHO PROSTŘEDÍ PRO SNADNÉ ZÍSKÁNÍ HLASIVKOVÝCH PULZŮ

Výše popsané naprogramované metody umožňují získat z řečového signálu průběh popisující proud vydechovaného vzduchu z plic skrze hlasivkovou štěrbinu, v anglické literatuře označovaný jako „glottal flow“. Aby mohl být určitý výpočetní algoritmus snadně ovládán z hlediska rychlosti (nezasahování do kódu při každé změně vstupních dat a nastavení parametrů metody), ergonomičnosti (funkčně uspořádané vstupní ovládací prvky a výstupní grafické indikátory) a zpracování získaných dat, je vhodné pro jeho ovládání vytvořit grafické uživatelské prostředí (GUI).

Jelikož jsou výpočetní metody naprogramované v interaktivním programovacím prostředí MATLAB 2015a, tak je z hlediska praktičnosti a zachování stejného vývojového prostředí, GUI vytvořeno pomocí nástroje implementovaného v tomto programu. Spouštěcí soubor „*GUI.m*“ pro inicializaci GUI se musí otevřít a spustit v prostředí programu MATLAB.

4.1 Ukázka vytvořeného GUI

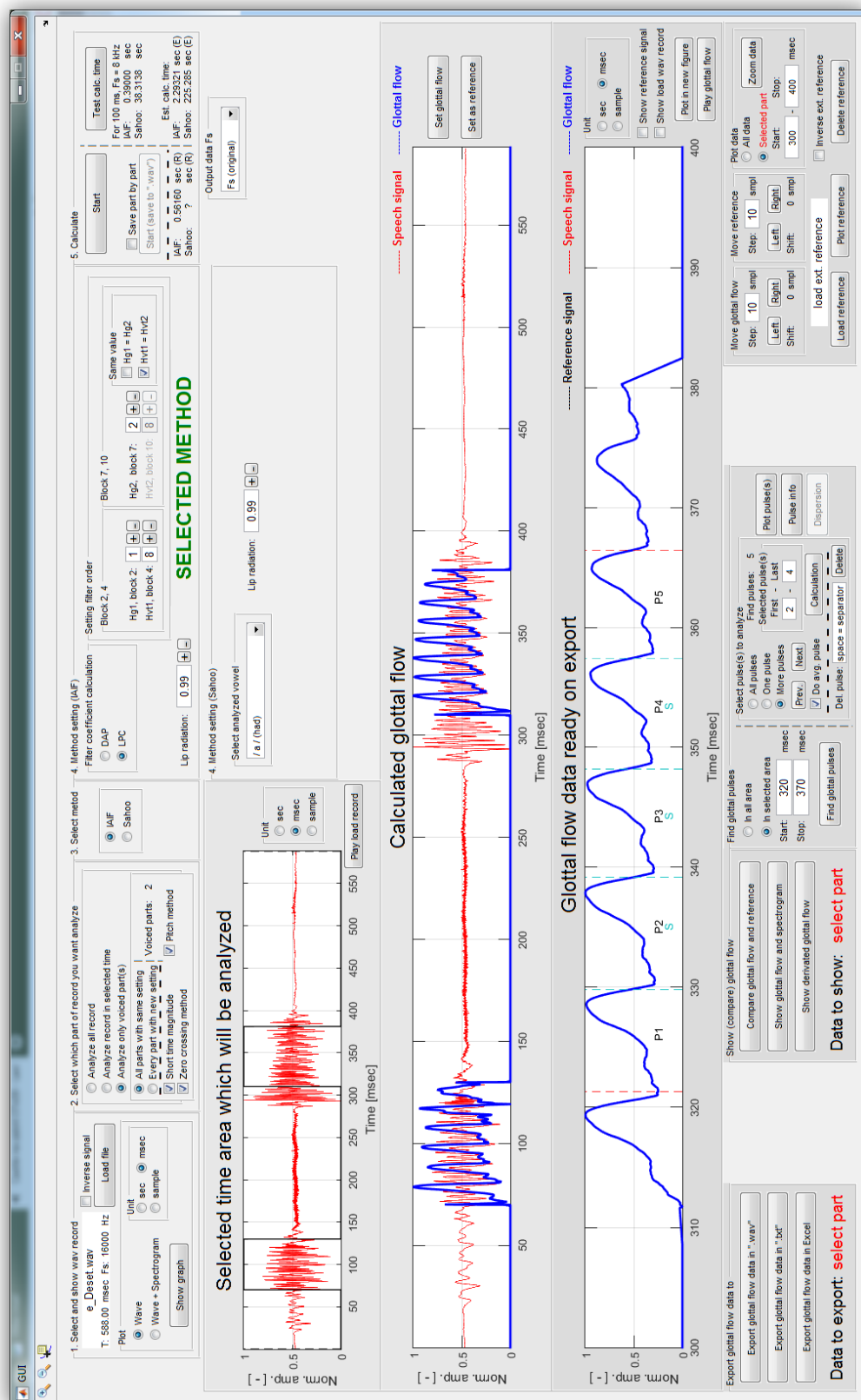
GUI je vytvořeno a optimalizováno pro monitory s obrazovým rozlišením Full HD (1920x1080 px), při použití na obrazovkách s menším rozlišením, nemusí být viděno celé okno programu, což zneprůjemňuje jeho ovládání.

V GUI je možné provádět základní nastavení zvolené výpočetní metody, o vhodnosti zvolených parametrů se můžeme přesvědčit ihned po vykonání výpočtu. Spočtená výstupní data ze zvolené techniky pro určení hlasivkových pulzů jsou ihned převedena do grafického průběhu vývoje pulzů v časové oblasti, kde můžeme zkontrolovat jejich tvar a v případě nespokojenosti s dosaženými výsledky změnit nastavení výpočetní metody.

Získaná výstupní data hlasivkových pulzů se mohou porovnat s referenčním signálem. Ten může být do programu nahrán interně nebo externě, jako interně nahraná reference se označuje signál vypočtený dříve s jiným nastavením metody, případně použití jiné výpočetní metody. Externě nahraná reference je jakýkoliv signál ve formátu „.WAV“ (dříve vypočtená a následně uložená data, případně data získaná z jiných zdrojů).

Důležitou vlastností programu, je možnost exportování získaných data do souboru z důvodu jejich zachování pro jejich pozdější analyzování. Ukládání je možné provést do zvukového souboru „.WAV“, textového souboru „.TXT“ a tabulkového programu Excel. Data je možné uložit do souboru ihned, jakmile jsme spokojeni s tvary hlasivkových pulzů, případně aktivovat možnost průběžného ukládání pulzů vždy po výpočtu daného znělého řečového úseku do souboru „.WAV“.

V následujících podkapitolách budou popsány jednotlivé bloky GUI, s ohledem na jejich význam, ovládací a výstupní prvky.



Obrázek 4.1 Vytvořené grafické uživatelské prostředí pro snadné získání hlasivkových pulzů při zpracovávání krátké zvukové nahrávky.

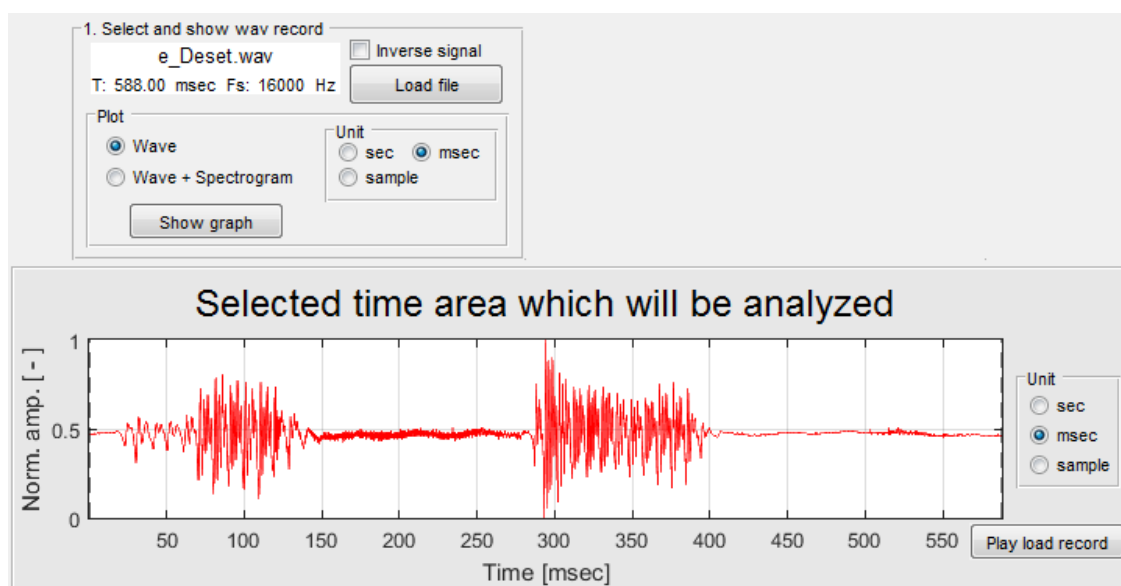
4.1.1 GUI, 1. blok: načtení řečového signálu a ukázání jeho časového průběhu

Vstupní data musí být do programu vkládána ve zvukovém formátu “.WAV“. Po výběru požadovaného řečového signálu pomocí tlačítka „Load file“ je zjištěna jeho délka, vzorkovací frekvence a provedeno vykreslení do grafu pod blokem 1. Pokud má signál vzorkovací frekvenci vyšší než 48 kHz, provede se jeho převzorkování na hodnotu 48 kHz a zobrazí se text „48k Re Hz“.

Pokud bylo po nahrání řečového signálu zjištěno, že je převrácený, stačí zaškrtnout volbu „Inverse signal“ a dojde k jeho otočení.

Je možné si zobrazit samostatný průběh řeči v novém okně nebo si k řečovému průběhu nechat spočítat a vykreslit i odpovídající spektrogram. Časové jednotky, ve kterých si chceme nechat vykreslit grafy, se nastavují hned vedle (napravo). Po stisku tlačítka „Show graph“ dojde k zobrazení požadovaného grafu se zvolenými jednotkami na časové ose.

Po úspěšném nahrání signálu došlo k jeho okamžitému zobrazení (červenou barvou) v grafu níže se jménem „Selected time area which will be analyzed“. Na pravé straně grafu se nalézá přepínač umožňující uživateli zvolit, jaké chce jednotky na časové ose. V pravém dolním rohu se nalézá tlačítko „Play load record“, po jeho stisknutí dojde k přehrání průběhu vykresleného v tomto grafu. Data zobrazovaná v grafu jsou ovládána také z bloku 2, viz následující kapitola.



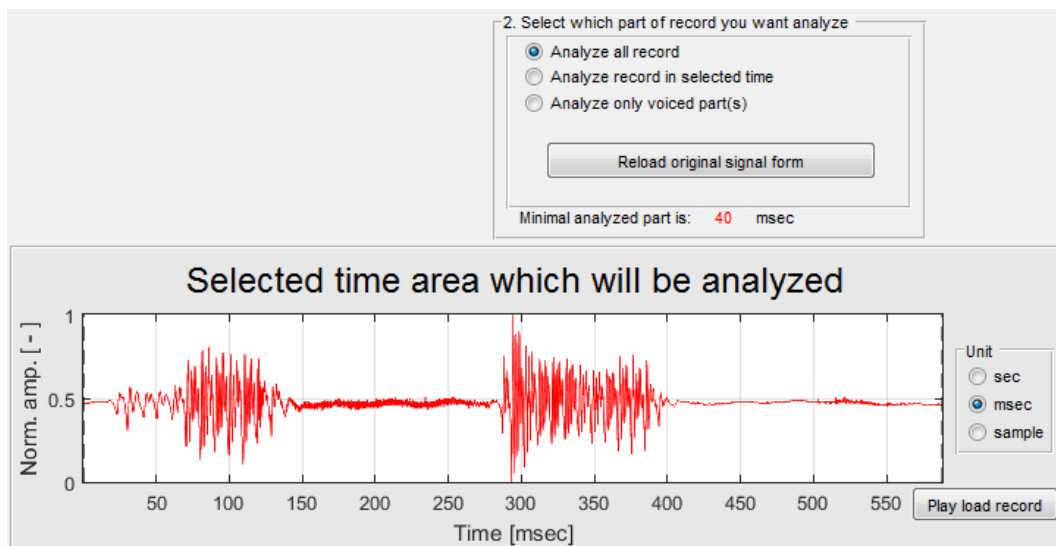
Obrázek 4.2 GUI, 1. blok: načtení řečového signálu a ukázání jeho časového průběhu.

4.1.2 GUI, 2. blok: výběr části signálu jaká má být analyzována

Je na výběr ze tří možností, jaká část signálu bude analyzována:

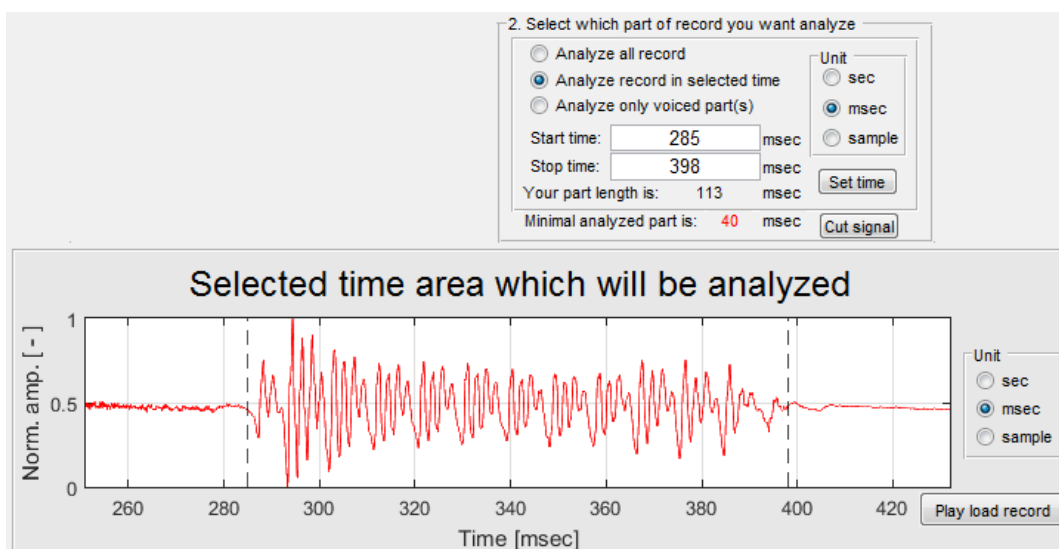
- Analyzovat celou nahrávku (Analyze all record)
- Analyzovat nahrávku ve vybraném čase (Analyze record in selected time)
- Analyzovat pouze znělé části (Analyze only voiced part(s))

První možnost „*Analyze all record*“ je nejjednodušší způsob jak má být signál analyzován, vezme se celá jeho část, na kterou se aplikuje vybraná výpočetní metoda. Tuto možnost výběru je možné vidět na Obrázku 4.3, dále je zde tlačítko „*Reload original signal form*“ to po stisku provede znovu načtení původního nahraného signálu.



Obrázek 4.3 GUI, 2. blok: část pro výběr celého signálu.

Druhá možnost „*Analyze record in selected time*“ je způsob kdy se z celého signálu vybere pouze požadovaný časový úsek, na něhož se následně aplikuje vybraná výpočetní metoda. Počáteční a koncový čas pro výběr části může být zadán v požadovaných časových jednotkách, potvrzení výběru je provedeno stiskem „*Set time*“ tím dojde k okopírování části signálu uprostřed ohraničené oblasti (čárkované černé čáry) které bude následně analyzována. Tlačítkem „*Cut signal*“ se původní řečový signál ořízne a zbyde v něm pouze část zobrazená mezi ohraničenou oblastí.



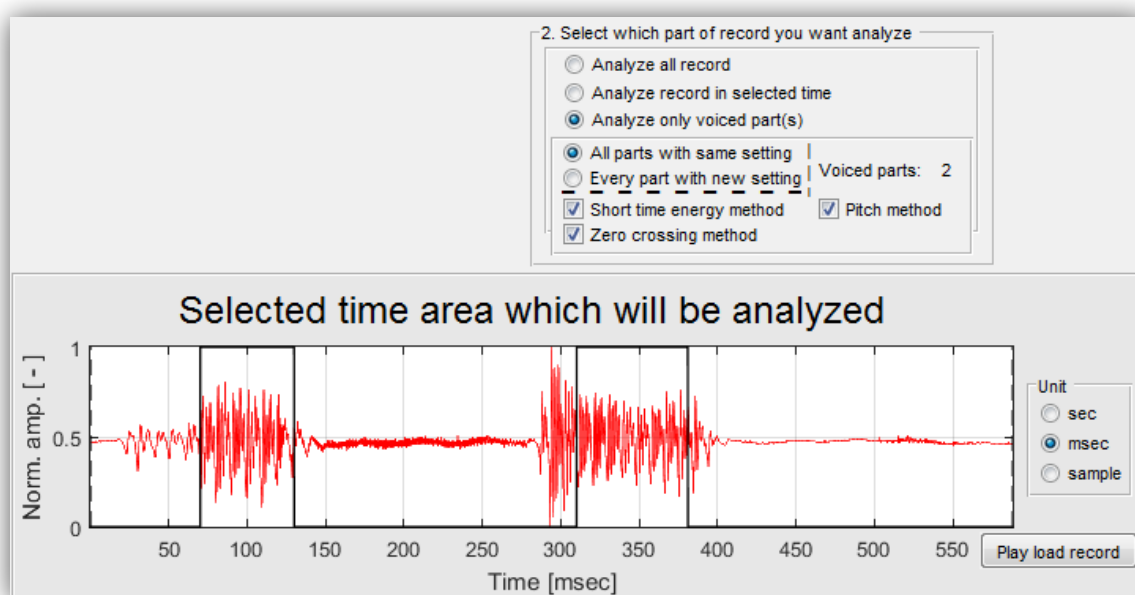
Obrázek 4.4 GUI, 2. blok: část pro výběr požadovaného časového úseku z celého signálu.

Třetí možnost „*Analyze only voiced part(s)*“ ihned po jejím výběru dojde k rozběhnutí výpočtu sloužícího pro detekování znělých částí řeči v celém signálu, zjištěný počet těchto částí je napsán v pravém horním rohu tohoto „podbloku“. Jejich zjištění se provádí pomocí tří metod (viz příloha):

- Metoda krátkodobé intenzity (Short time magnitude)
- Metoda detekování kmitočtu hlasivkových pulzů (Pitch method)
- Metoda počtu průchodů nulou (Zero crossing method)

Při základním nastavení jsou všechny metody aktivní, ale pomocí zaškrtnutí políčka si můžeme vybrat, zda chceme aktivní všechny či jen některé. Začátek a konec úseku znělosti je zjištěn pomocí průniku, kdy jsou kladně vyhodnoceny hodnoty znělosti ze všech použitých postupů. Stanovené znělé části jsou zobrazeny do grafu (černou barvou (1 = znělá část, 0 = neznělá část)) s řečovým signálem a také použity jako vstupy do vybrané výpočetní metody.

Pokud si z bloku dostupného u třetí možnosti vybereme „*All parts with same setting*“ budou všechny znělé části zpracovány se stejným nastavením výpočetní metody. Při zvolení „*Every part with new setting*“ se každý znělý úsek bude analyzovat zvlášť (nový úsek, nová možnost nastavením výpočetní metody).



Obrázek 4.5 GUI, 2. blok: detekování znělých částí řeči z řečového signálu.

4.1.3 GUI, 3. blok: výběr výpočetní metody pro získání hlasivkových pulzů

V tomto bloku programu se vybírá, jaká bude použita výpočetní metoda pro získání hlasivkových pulzů. Je zde na výběr ze dvou metod (popsaných v předchozích kapitolách):

- IAIF
- Sahoo



Obrázek 4.6 GUI, 3. blok: výběr výpočetní metody pro získání hlasivkových pulzů.

Pokud byla zvolena možnost IAIF je ve 4. bloku pro metodu IAIF zeleně zobrazen nápis „SELECTED METHOD“ výrazně informující uživatele o vybrané výpočetní metodě.

Pokud byla zvolena možnost Sahoo je provedena obdobná věc jako u IAIF.

4.1.4 GUI, 4. blok I.: nastavení základních hodnot pro metodu IAIF

V tomto bloku programu, který je rozdělen na několik menších bloků (boxů) se nastavují základní hodnoty pro výpočet metodou IAIF.

Blok „**Filter coefficient calculation**“ slouží k určení, jakým způsobem budou vypočítávány koeficienty filtrů. Je na výběr ze dvou možností:

- DAP (Discrete all-pole modeling)
- LPC (Linear predictive coding)

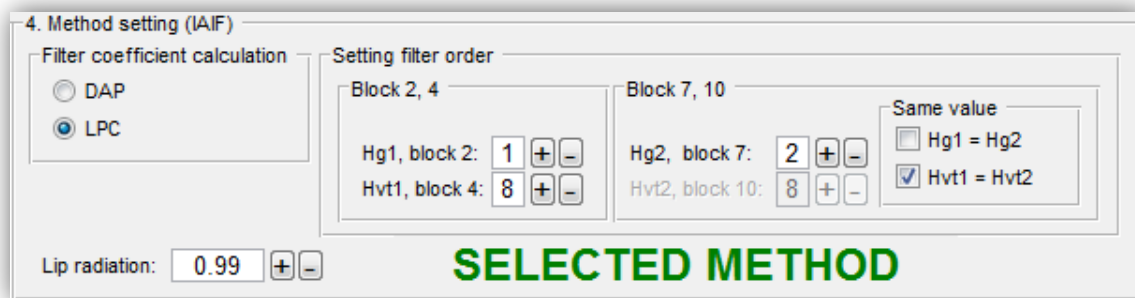
Získání koeficientů pro filtr pomocí DAP je popsáno v příloze. Získání koeficientů pro filtr LPC je řešeno pomocí implementované vnitřní funkce v programu MATLAB.

Blok „**Setting filter order**“ v sobě sdružuje dva boxy sloužící k nastavení řádů filtrů pro metodu IAIF a jejich nastavení je popsáno v kapitole 2.1.

První box s označením „**Block 2, 4**“ odpovídá svým názvem popisu bloků z vývojového diagramu pro tento algoritmus (viz Obrázek 2.2). Nastavení řádu filtru se může provést zapsáním hodnoty do jeho editačního pole nebo kliknutím na tlačítko „+“ nebo „-“ čímž dojde k navýšení či snížení hodnoty v editačním poli.

Druhý box s označením „**Block 7, 10**“ má stejné vlastnosti jako předchozí box a navíc v sobě ještě obsahuje box „**Same value**“. Ten při zaškrtnutí „ $Hg1 = Hg2$ “ nebo „ $Hg1 = Hg2$ “ slouží k udržení stejné hodnoty řádu filtru jako v prvním boxu (automaticky překopíruje hodnotu řádu filtru z prvního boxu do druhého).

Poslední hodnota, která může být editována je „**Lip radiation**“ a slouží k nastavení netěsnosti ρ u integrátoru s netěsností. Ten se používá pro eliminování vlivu hlasivek na vstupní řečový signál a také na eliminaci efektu vyzařování skrze rty.



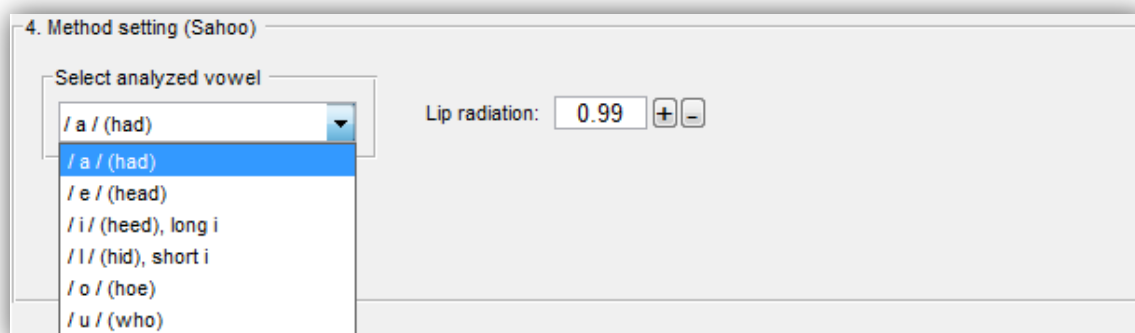
Obrázek 4.7 GUI, 4. blok I.: nastavení základních hodnot pro metodu IAIF.

4.1.5 GUI, 4. blok II.: nastavení základních hodnot pro metodu Sahoo

V tomto bloku programu se provádí nastavení parametrů pro získání hlasivkových pulzů pomocí metody Sahoo.

V rozkládacím menu „**Select analyzed vowel**“ se musí nastavit koeficienty válcového modelu hlasového traktu pro právě analyzovanou samohlásku dostupné z [31].

Následně stejně jako v předchozí metodě, může být editována hodnota „**Lip radiation**“ sloužící k nastavení netěsnosti ρ u integrátoru s netěsností. Ten se používá na eliminaci efektu vyzařování skrze rty.



Obrázek 4.8 GUI, 4. blok II.: nastavení základních hodnot pro metodu Sahoo.

4.1.6 GUI, 5. blok I.: zahájení výpočtu pomocí zvolené a nakonfigurované metody

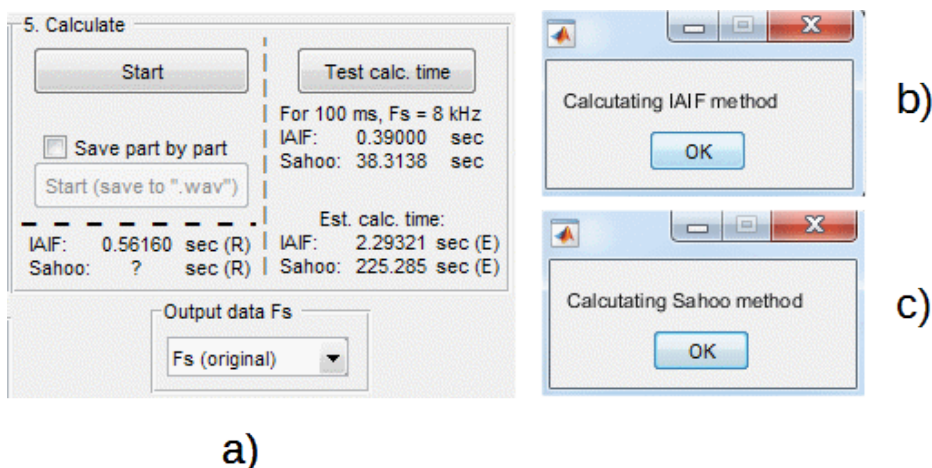
Finální blok obstarávající spuštění výpočetní dříve zvolené a následně vhodně nakonfigurované výpočetní metody.

V rozkládacím menu „**Output data Fs**“ se určuje, zda výstupní data z vybrané metody mají být ponechána na stejné vzorkovací frekvenci jako při vstupu do metody nebo převzorkovány na jinou zvolenou vzorkovací frekvenci. Toto nastavení musí být provedeno před stiskem tlačítka „*Start*“.

Stiskem tlačítka „*Start*“ dojde ke spuštění dříve nakonfigurované metody a zobrazí se box informující o probíhajícím výpočtu (viz Obrázek 4.8 **b**), **c**). Po skončení výpočtu je do 5. bloku vlevo dole, k použité metodě s označením (**R**) - (Real calc. time) zapsán reálný výpočetní čas, zmizí informační box a získaná data se vykreslí do grafu s názvem „**Calculated glottal flow**“.

Stiskem tlačítka „*Test calc. time*“ dojde k inicializaci testovací výpočtové sekvence pro obě metody. Následně je automaticky proveden odhad potřebného výpočetního času pro výpočet hlasivkových pulzů z celého nahraného řečového signálu. Odhadovaný čas je zapsán do 5. bloku vpravo dole, k použité metodě s označením (**E**) - (Estimated calc. time).

Tlačítko „*Start (save to ".wav")*“ se aktivuje po zaškrtnutí možnosti „*Save part by part*“ a to pouze pokud je vybrána možnost analyzování znělých úseků. Po jeho stisku se rozběhne dříve nakonfigurovaná metoda IAIF a po každém vypočteném úseku jsou získaná data uložena do zvukového souboru „.WAV“.



Obrázek 4.9 GUI, 5. blok I.: **a)** blok obstarávající spuštění výpočtu pomocí zvolené a nakonfigurované metody; **b)** zpráva zobrazená při výpočtu metodou IAIF; **c)** zpráva zobrazená při výpočtu metodou Sahoo.

4.1.7 GUI, 5. blok II.: výpočet hlasivkových pulzů při volbě „*Every part with new setting*“ z 2. bloku

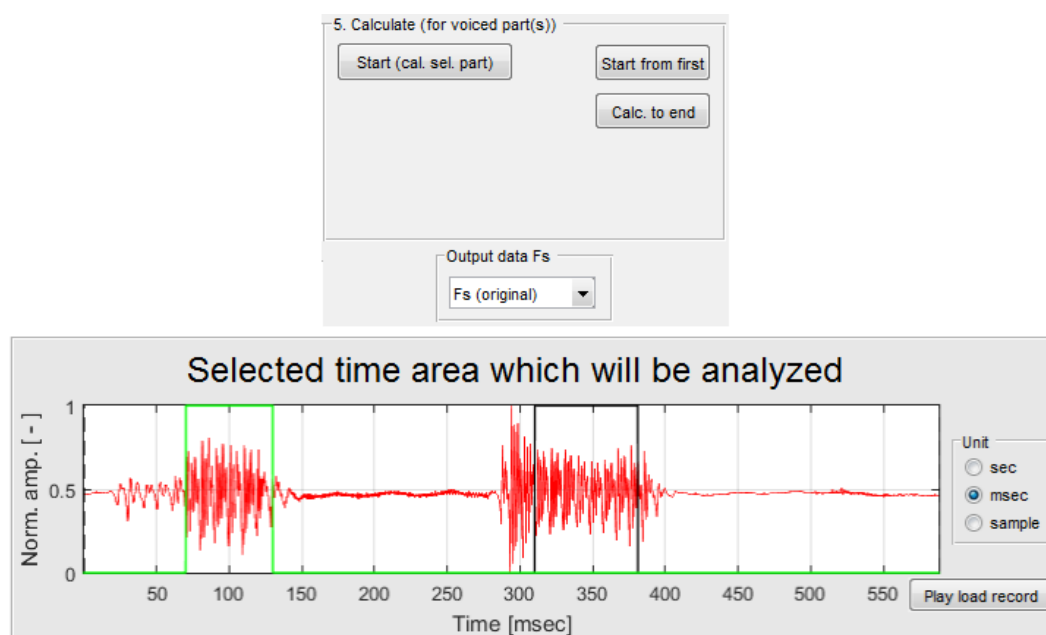
Finální blok obstarávající spuštění výpočetní dříve zvolené a následně vhodně nakonfigurované výpočetní metody pro znělé úseky pokud je chceme zpracovávat část

po části. Dostupná pouze při výběru možnosti „*Every part with new setting*“ z 2. bloku.

Tlačítko „*Start (cal. sel. part)*“ slouží ke spuštění dříve nakonfigurované metody se vstupními daty prvního pulzu, v grafu „**Selected record wav area which will be analyzed**“ je tento i každý další právě analyzovaný úsek ohraničen obdélníkem se zelenými hranami. Pokud je tvar pulzu vyhovující, potvrdí se stiskem tlačítka „*Set glottal flow*“ z bloku s grafem „**Calculated glottal flow**“.

Tlačítko „*Start from first*“ slouží k restartování výpočtu tj. přesunutí na první znělý úsek.

Po stisku tlačítka „*Calc. to end*“ se aktivuje smyčka, která vypočítá všechny zbývající znělé úseky s právě vybranou metodou a jejím současným nastavením.



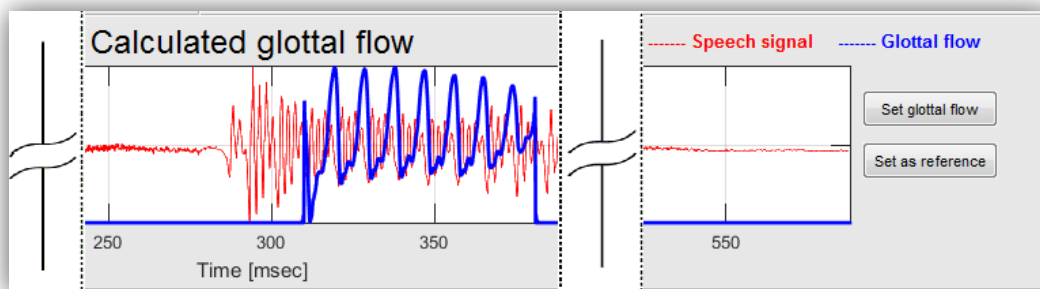
Obrázek 4.10 GUI, 5. blok II.: blok pro výpočet hlasivkových pulzů při volbě „*Every part with new setting*“ z 2. bloku.

4.1.8 GUI, blok: graf „**Calculated glottal flow**“

Blok s grafem „**Calculated glottal flow**“ slouží k zobrazení právě vypočtených dat průběhu hlasivkových pulzů.

Tlačítkem „*Set glottal flow*“ se právě zobrazená výstupní data z vybrané metody přesunou do grafu „**Glottal flow data ready on export**“ a jsou označena jako glottal flow.

Stiskem tlačítka „*Set as reference*“ se právě zobrazená výstupní data z vybrané metody přesunou do „**Glottal flow data ready on export**“ a jsou označena jako referenční signál.



Obrázek 4.11 GUI, blok: graf „Calculated glottal flow“ jeho význam a ovládací prvky.

4.1.9 GUI, blok: graf „Glottal flow data ready on export“

Blok s grafem „Glottal flow data ready on export“ slouží k zobrazení hlasivkových pulzů před jejich exportováním do souboru. Je možné si nastavit, v jakých časových jednotkách má být graf vykreslen.

Zaškrtnutí „Show reference signal“ po aktivování zobrazí v grafu referenční signál.

Zaškrtnutí „Show load wav record“ po aktivaci zobrazí v grafu nahraný řečový signál.

Tlačítko „Plot in new figure“ po stisknutí vykreslí všechny data aktuálně zobrazené v „Glottal flow data ready on export“ do nového okna.

Tlačítko „Play glottal flow“ po stisknutí přehraje (pustí do výstupu zvukové karty) celý zobrazený průběh signálu hlasivkových pulzů.

Blok „Move glottal flow“ slouží k posunu (shift) průběhu hlasivkových pulzů po časové ose doleva či doprava. V editačním okně je možné si navolit krok, po kolika vzorcích se bude signál posouvat. U spodu tohoto bloku je možné vidět, o kolik vzorků byl signál posunut.

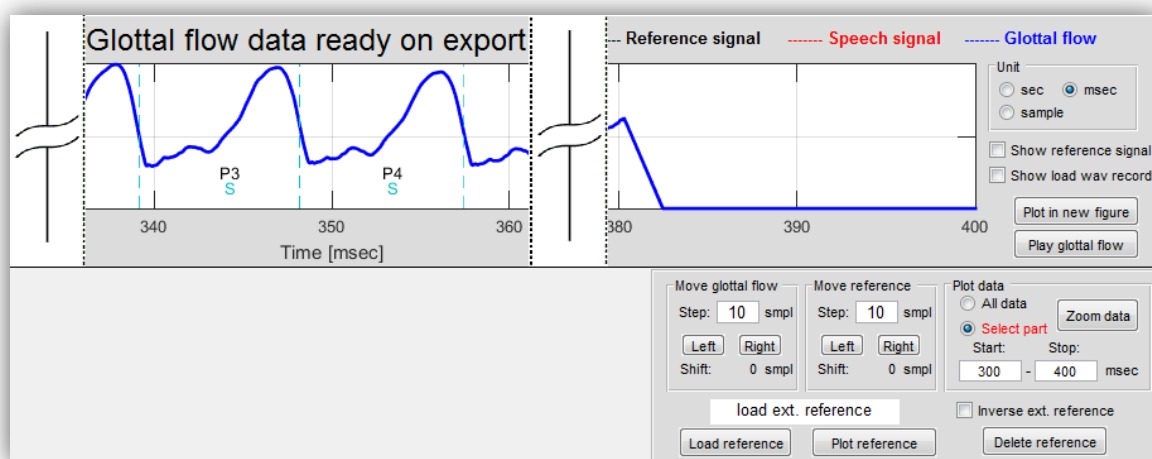
Blok „Move reference“ slouží k posunu (shift) průběhu referenčního signálu po časové ose doleva či doprava. Ovládání je stejné jako v předchozím případě.

Blok „Plot data“ slouží k přiblížení vybraného úseku na grafu, je možné si zobrazit:

- Všechny data (All data)
- Vybranou část (Selected part)

Po vybrání možnosti „All data“ je do grafu vykreslen celý signál průběhu hlasivkových pulzů. Zvolením možnosti „Selected part“ a následným stiskem tlačítka „Zoom data“ se v grafu zobrazí data v rozmezí, jaké bylo určeno pomocí „Start:“ a „Stop:“ hodnoty přečtené z editačního boxu. Časovou jednotku v jakých mají být zadány určující meze zobrazených dat, nalezneme na pravé straně stop boxu. Závisí na jednotkách, v jakých byl vykreslen celý graf. Při stisku „Zoom data“ dojde také k zčervenání možnosti „Select part“ aby bylo vidět, že při exportování, porovnávání či kreslení spektrogramu bude využito dat pouze ze zvoleného úseku!

Tlačítko „*Load reference*“ slouží k načtení reference z externího souboru ve formátu „.WAV“. Následně je možné si nahranou referenci zkontrolovat, zda je v pořádku v novém okně po stisku tlačítka „*Plot reference*“. Pokud bylo shledáno, že je potřeba externě nahranou referenci převrátit, použijeme zaškrtnuté tlačítko „*Inverse ext. reference*“. Po stisku tlačítka „*Delete reference*“ dojde k vymazání dat referenčního signálu.



Obrázek 4.12 GUI, blok: graf „**Glottal flow data ready on export**“ jeho význam a ovládací prvky.

4.1.10 GUI, blok: Find glottal pulses

Blok „**Find glottal pulses**“ slouží k nalezení a lokalizování jednotlivých hlasivkových pulzů z celého signálu popisující vývoj hlasivkových pulzů v čase. Separované pulzy s pořadovým číslem jsou zakresleny do grafu pomocí červených svislých přerušovaných čar. Jednotlivé pulzy mohou být lokalizovány:

- V celé oblasti (In all area)
- Ve vybrané oblasti (In selected area)

Při vybrání možnosti „*In all area*“ a následným potvrzením pomocí tlačítka „*Find glottal pulses*“ bude prohledána oblast celého signálu a lokalizuje všechny hlasivkové pulzy. Zvolením „*In selected area*“ a znovu následných potvrzením tlačítkem „*Find glottal pulses*“ dojde k lokalizování pulzů pouze v oblasti vymezující se pomocí „*Start*“ a „*Stop*“ hodnoty načtené z editačního boxu. Časová jednotka, ve které se vymezuje oblast je zvolena stejně jako v předchozí kapitole (bloku).

Dále se zde nalézá blok s popisem „**Select pulse(s) to analyze**“ zabývající se výběrem jednotlivých pulzů získaných v předchozím bloku, jejich počet je zobrazen v pravém horním rohu tohoto bloku. Tyto pulzy je možné ještě vybrat a zobrazit pomocí třech kritérií:

- Všechny pulzy (All pulses)
- Jeden pulz (One pulse)
- Více pulzů (More pulses)

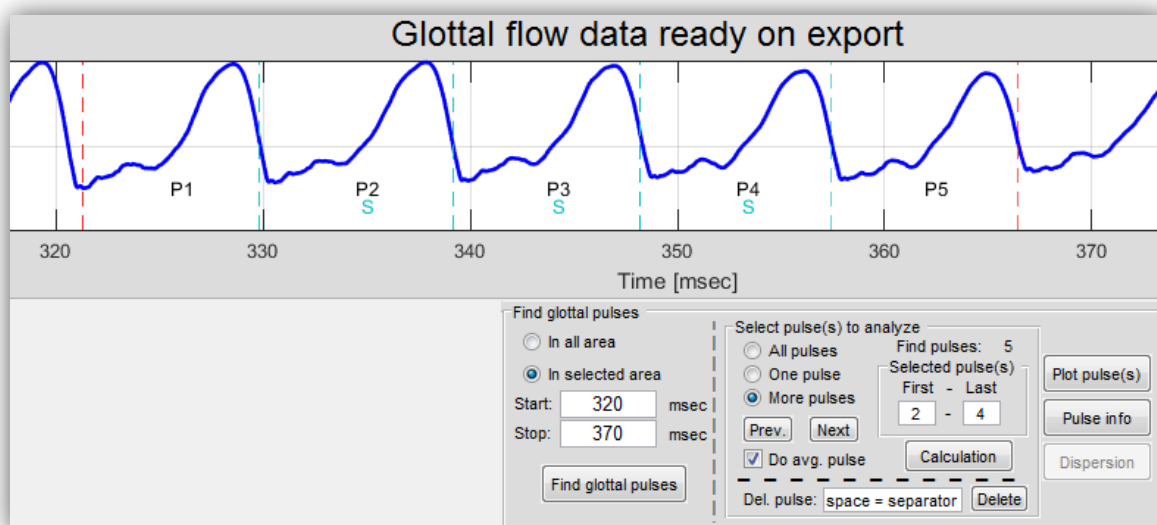
Při způsobu výběru „*All pulses*“ a „*More pulses*“ může být aktivována možnost „*Do avg. pulse*“ sloužící k vypočtení průměrného tvar hlasivkového pulzu ze všech detekovaných pulzů. Možnost „*One pulse*“ vybere pouze jeden pulz, mezi jednotlivými pulzy se dá přecházet pomocí tlačítek „*Prev.*“ a „*Next*“ nebo zadat přímo číslo požadovaného pulzu do editačního boxu. Při možnosti „*More pulse*“ se do editačního boxu „*First*“ a „*Last*“ zadává číslo prvního a posledního pulzu který má být podroben analýze. Po stisku tlačítka „*Calculation*“ dojde k vybrání pouze zvolených pulzů. Vybrané pulzy mají pod pořadovým číslem modré **S** (Selected) a jsou v grafu označeny pomocí modrých svislých přerušovaných čar.

V bloku „**Select pulse(s) to analyze**“ úplně dole je možnost odstranění nalezeného pulzu ze seznamu detekovaných pulzů například z důvodu špatného určení pulzu. Stačí zadat pořadové číslo pulzu, pokud je pulzů více, oddělují se mezerou a po stisknutí tlačítka „*Delete*“ je záznam o pulzu vymazán.

Tlačítko „*Plot pulse(s)*“ otevře nové okno a vykreslí všechny vybrané pulzy přes sebe.

Tlačítko „*Pulse info*“ otevře nové okno a zobrazí zvolený pulz a vypíše o něm zjištěné informace.

Tlačítko „*Disperzion*“ otevře nové okno a zobrazí průměrný pulz doplněný o směrodatnou odchylku.



Obrázek 4.13 GUI, blok: Find glottal pulses.

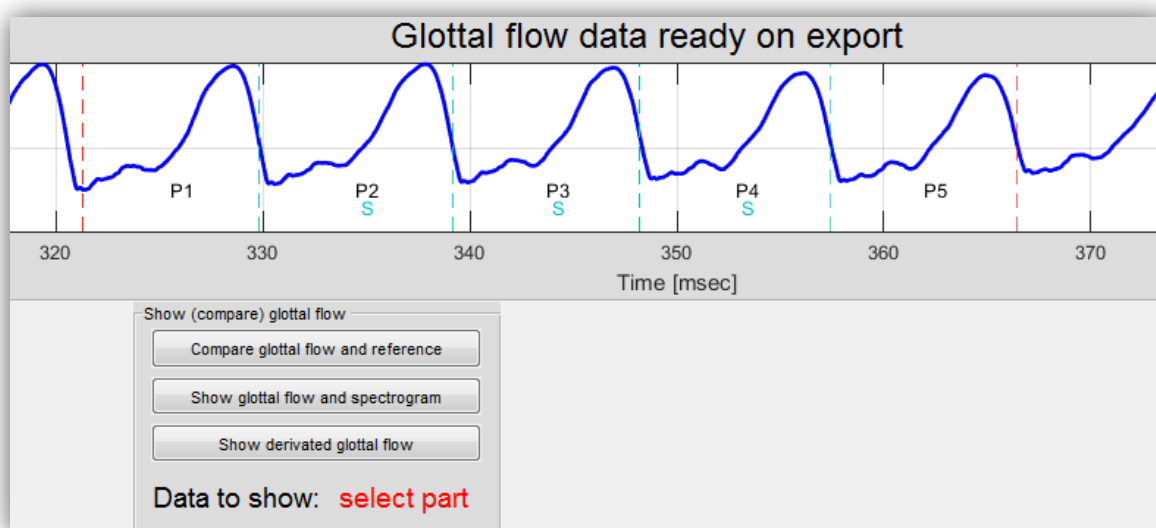
4.1.11 GUI, blok: Show (compare) glottal flow

Blok „**Show (Compare) glottal flow**“ obsahuje tři tlačítka. Ve spodní části bloku je zobrazena informace, zda je vybrán celý signál nebo pouze jeho část.

Tlačítko „*Compare glottal flow and reference*“ otevře nové okno, ve kterém je ukázáno porovnání hlasivkového pulzu a referenčního signálu.

Tlačítko „*Show glottal flow and spectrogram*“ otevře nové okno ve kterém je vidět průběh hlasivkových pulzů v čase a k tomu odpovídající spektrogram.

Tlačítko „*Show derivated glottal flow*“ otevře nové okno, ve kterém bude zobrazen derivovaný průběh hlasivkového pulzu.



Obrázek 4.14 GUI, blok: Show (compare) glottal flow.

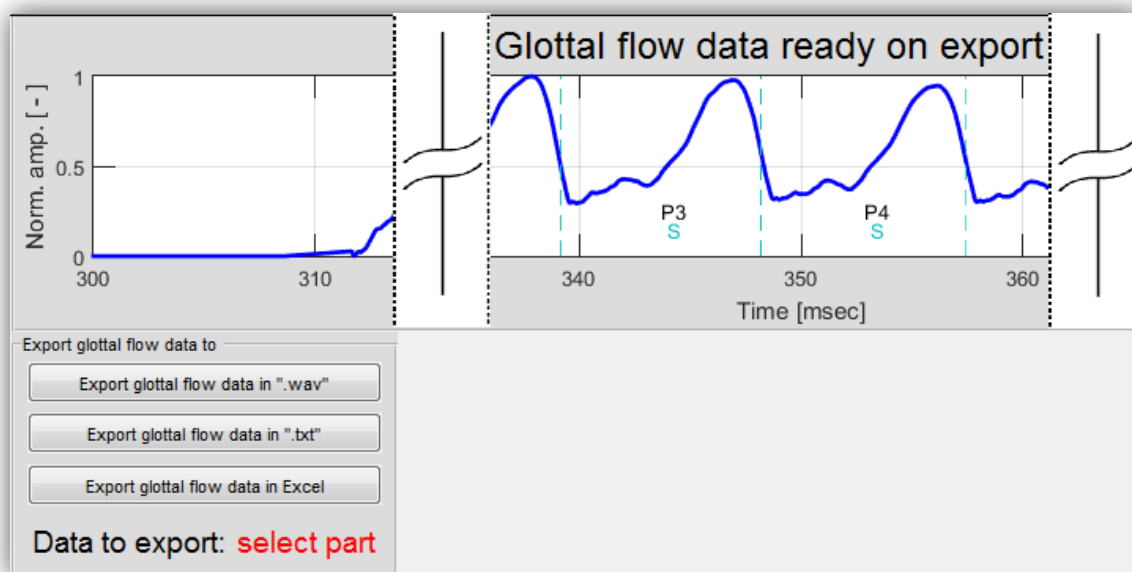
4.1.12 GUI, blok: Export glottal flow data to

Blok „**Export glottal flow data to**“ slouží k vyexportování (uložení) získaných dat tvarů hlasivkových pulzů do souboru. Ve spodní části bloku je zobrazena informace, zda je vybrán celý signál nebo pouze jeho část.

Tlačítko „*Export glottal flow data in “.wav”*“ exportuje průběh hlasivkových pulzů do zvukového souboru “.WAV”.

Tlačítko „*Export glottal flow data in “.txt”*“ exportuje data průběhu hlasivkových pulzů do textového souboru.

Tlačítko „*Export glottal flow data in Excel*“ exportuje data do excelového souboru.



Obrázek 4.15 GUI, blok: Export glottal flow data to.

4.2 Podrobnější popis určitých funkcí pro bloky GUI

4.2.1 GUI, blok: graf „Calculated glottal flow“

Tlačítko „*Set glottal flow*“ při vybrané volbě analyzování pouze znělých úseků zajišťuje eliminaci skokové změny hodnoty (vytvoření špičky) z nuly na hodnotu pulzu získanou výpočtem. Vyrušení špičky je provedeno lineárním náběhem (případně sestupem) z nulové hodnoty na hodnotu pulzu získanou výpočtem. Minimální doba trvání lineárního náběhu (sestupu) je 2 ms a maximální 4 ms.

Ukázka lineárního náběhu je vidět na Obrázku 4.15 a lineárního sestup na Obrázku 4.12.

4.2.2 GUI, blok: Find glottal pulses

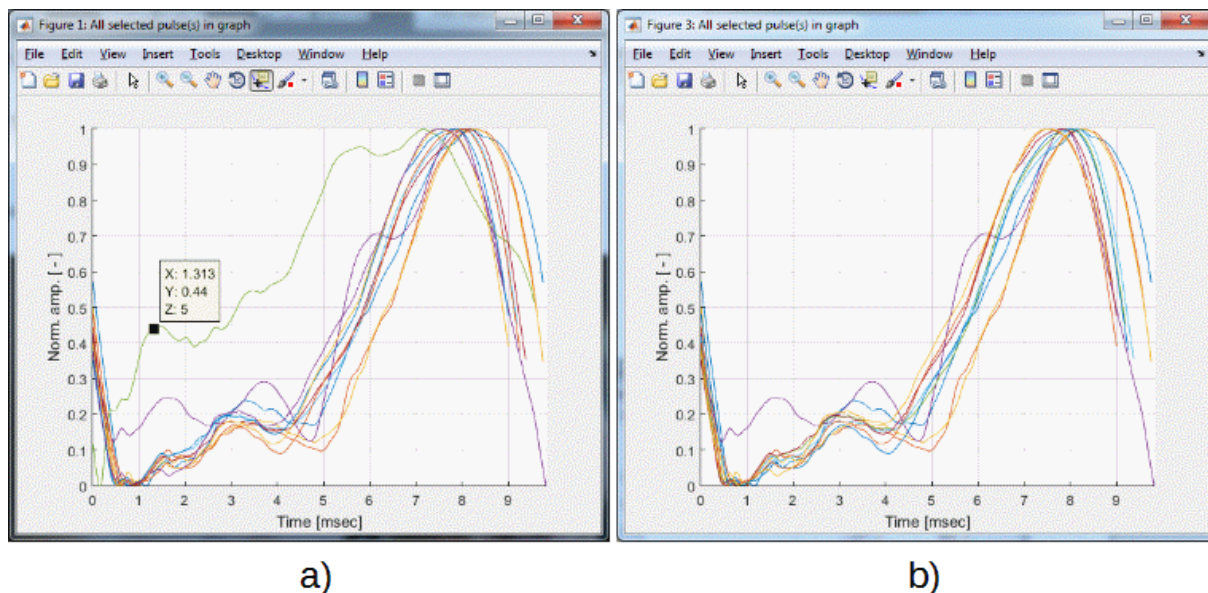
Lokalizování jednotlivých hlasivkových pulzů je provedeno následovně:

- 1) Provedení derivace na celém průběhu signálu hlasivkových pulzů.
- 2) Najít lokální minima v derivovaném průběhu, odpovídající okamžikům uzavření hlasivek (GCI).
- 3) Uložit pozice GCI a následně je vyznačit do grafu.

Tlačítko „*Plot pulse(s)*“ otevře nové okno a vykreslí všechny vybrané nalezené pulzy přes sebe. Po vybrání možnosti z horní lišty „*Data cursor*“ a následném kliknutí na průběh hlasivkového pulzu, se zobrazí jeho hodnoty pro osu X a Y , Z hodnota udává pořadové číslo pulzu, viz Obrázek 4.16 a).

Tlačítko „*Delete*“ přečte jednotlivá čísla pulzů (oddělené mezerami) zapsané v editačním boxu a odstraní je ze seznamu všech nalezených pulzů v zadané oblasti.

Typická ukázka pro užití tohoto tlačítka je na Obrázku 4.16 **a)** kdy zelený puls vybočuje svým tvarem (pravděpodobně byl špatně odhadnut jeho tvar) z množiny více pulzů a způsoboval by nepřesnosti při výpočtech na jednotlivých pulzech v tomto bloku, proto byl odstraněn ze seznamu nalezených pulzů, viz Obrázek 4.16 **b)**.



Obrázek 4.16 Okno zobrazující: **a)** všechny nalezené pulzy, **b)** odstraněný chybný pulz číslo 5.

Tlačítko „Pulse info“ otevře nové okno, zobrazí zvolený pulz a vypíše o něm zjištěné informace:

- Fáze otevírání (Opening phase)
- Fáze uzavírání (Closing phase)
- Celková doba (Total time)
- Frekvence (Frequency)
- Špičatost (Kurtosis)
- Šikmost (Skewness)

Jako začátek fáze otevírání byl zvolen okamžik, kdy je dosaženo minimální hodnoty na levé straně pulzu, konec fáze otevírání nastává při dosažení maximální hodnoty pulzu (černá čárkovaná čára).

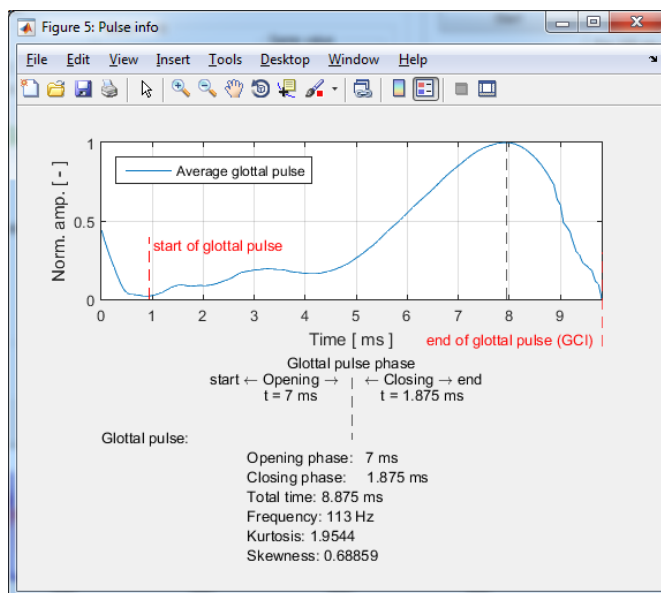
Fáze uzavírání trvá od maximální hodnoty pulzu ke GCI (konci pulzu).

Špičatost pulzu nabývá následujících hodnot:

- Normální rozdělení má hodnotu **3**
- Hodnota **větší než 3** značí kladnou špičatost (hodnoty jsou soustředěny kolem normálního rozdělení (střední hodnoty))
- Hodnota **menší než 3** značí zápornou špičatost (hodnoty jsou vzdálené od normálního rozdělení (střední hodnoty))

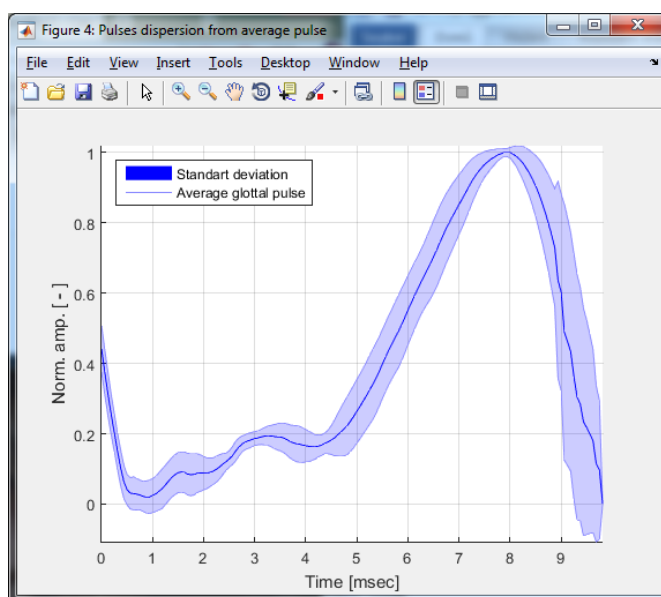
Šikmost pulzu nabývá následujících hodnot:

- Normální rozdělení má hodnotu **0**
- Záporná hodnota (většina hodnot se nachází nad střední hodnotou)
- Kladná hodnota (většina hodnot se nachází pod střední hodnotou)



Obrázek 4.17 Okno zobrazující informace o zvoleném pulzu.

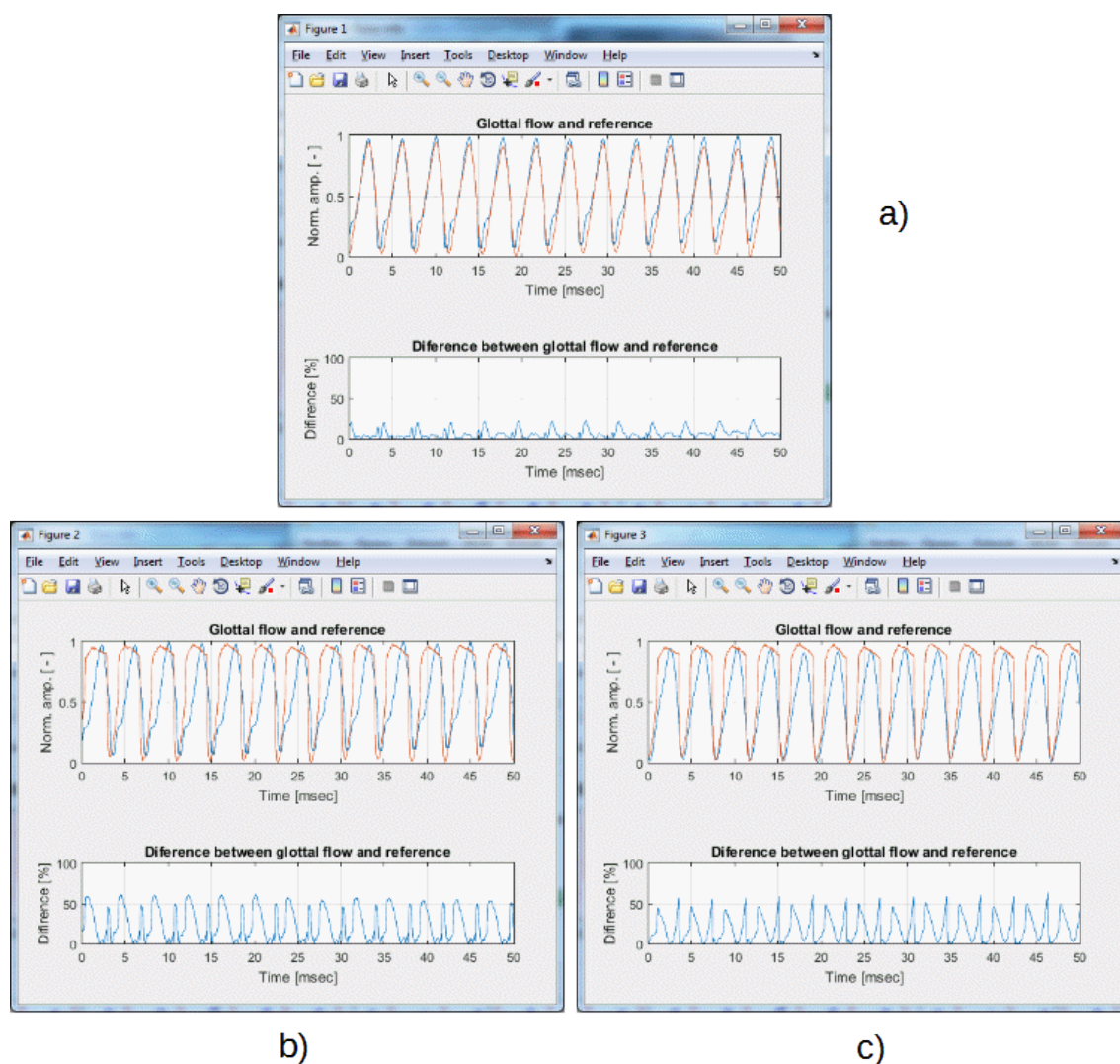
Tlačítko „Disperzion“ otevře nové okno a zobrazí průměrný pulz doplněný o směrodatnou odchylku. Hodnota směrodatné odchylky v jednom bodě je stejná (celá směrodatná odchylka, ne půl odchylky) nahoru i dolů od zvoleného bodu.



Obrázek 4.18 Okno zobrazující směrodatnou odchylku od průměrného pulzu.

4.2.3 GUI, blok: Show (compare) glottal flow

Tlačítko „Compare glottal flow and reference“ otevře nové okno, ve kterém je ukázáno porovnání hlasivkového pulzu a referenčního signálu. Jako referenční signál se může použít hlasivkový průběh vypočtený s jiným nastavením či metodou viz Obrázek 4.19 a) nebo průběh nahrát externě. Externě se může nahrát jakýkoliv průběh ve formátu „.WAV“, například záznam z EGG nahraný ve stejnou dobu jako nahrávka získaná přes mikrofon [32], viz Obrázek 4.19 b), c).



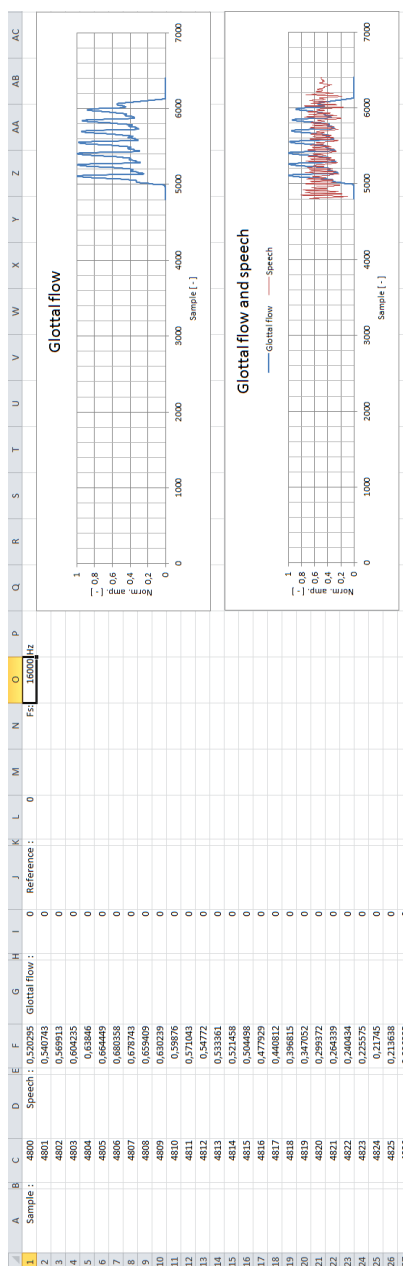
Obrázek 4.19 a) porovnání výstupu metody IAIF a Sahoo na stejný řečový signál, b) porovnání výstupu metody IAIF a dat z EGG, c) porovnání výstupu metody Sahoo a dat z EGG.

Tlačítko „Show glottal flow and spectrogram“ otevře nové okno ve kterém je vidět průběh hlasivkových pulzů v čase a k tomu odpovídající spektrogram. Pro rychlejší výpočet spektrogramu se vstupní průběh hlasivkových pulzů převzorkuje na $F_s = 8$ kHz, horní hranice zobrazeného kmitočtu na spektrogramu má maximální hodnotu 2 000 Hz.

4.2.4 GUI, blok: Export glottal flow data to

Tlačítko „*Export glottal flow data in .txt*“ exportuje data průběhu hlasivkových pulzů do textového souboru. V souboru je uvedena informace o použité vzorkovací frekvenci dat, průběh signálu hlasivkových pulzů a průběh řečového signálu v čase. Jako oddělovač mezi daty je použit středník.

Tlačítko „*Export glottal flow data in Excel*“ exportuje data do excelového souboru. V souboru je uvedena informace o použité vzorkovací frekvenci dat, průběh signálu hlasivkových pulzů, průběhu referenčního signálu a průběh řečového signálu v čase. Jednotlivé vzorky jsou seřazeny do sloupců, ze vzorků jsou ihned vytvořeny odpovídající grafy, viz Obrázek 4.20. Export dat do tohoto souboru trvá delší dobu.



Obrázek 4.20 Ukázka exportu dat do excelového souboru.

5 UKÁZKA VYUŽITÍ TVARU ZÍSKANÝCH HLASIVKOVÝCH PULZŮ

Pro otestování funkčnosti všech výše uvedených funkcí GUI, byly v programu analyzovány krátké řečové úseky, stejně tak i dlouhé řečové úseky.

Test zpracování krátkých úseků řeči byl proveden na samohlásce /a/ o délce 100 ms a slově „Deset“ obsahující dvě samohlásky /e/. Test zpracování dlouhých úseků řeči byl proveden na delších úsecích řečových signálů (řádově minimálně jednotky sekund), obsahující více znělých úseků.

Na vyzkoušení, zda program umí analyzovat nalezené hlasivkové pulzy, byl stažen balík obsahující zvukové nahrávky ze studentské soutěže „BIOSIGNAL CHALLENGE 2018“ [35] zabývající se detekcí Parkinsonovy choroby pomocí řečových promluv. Z celého balíku zvukových nahrávek bylo zpracováno 40 nahrávek, z toho 20 patřilo lidem ze zdravé kontrolní skupiny (HC), pro každého jedince jsou dostupné dvě nahrávky („a“, „b“) celkově tedy byly analyzovány hlasivkové pulzy od 10 zdravých lidí. Zbylých 20 nahrávek patřilo lidem s diagnostikovanou Parkinsonovou nemocí (PN), kde pro každého jedince jsou stejně jako v předchozím případě dostupné dvě nahrávky („a“, „b“).

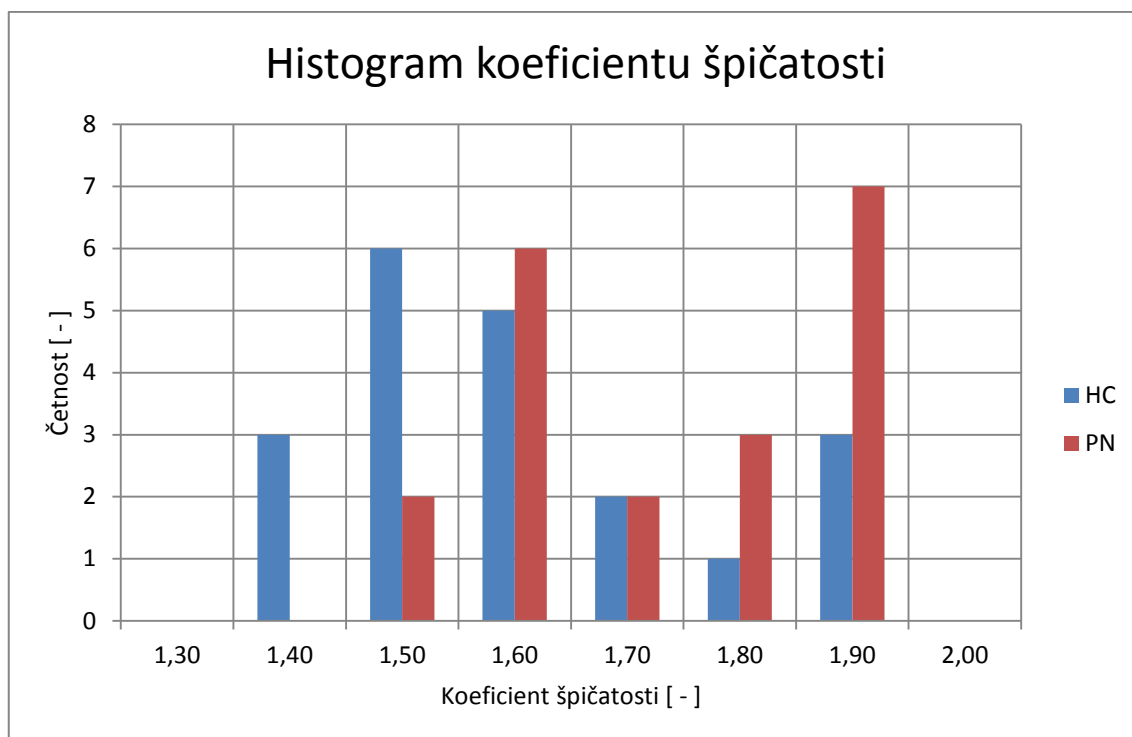
Výpočet průběhu hlasivkových pulzů z každé vybrané nahrávky je proveden pomocí metody IAIF, následně byly nalezeny jednotlivé hlasivkové pulzy a z těch jsou odebrány špatně detekované (případně vypočtené) pulzy. Po odebrání všech rušivých průběhů je vypočten ze všech jednotlivých pulzů průměrný pulz, z něhož byla zjištěna hodnota koeficientu špičatosti a šikmosti průměrného průběhu.

Subjekt	Špičatost	Šikmost
HC631a	1,52	0,15
HC631b	1,50	0,15
HC632a	1,38	0,05
HC632b	1,39	0,09
HC633a	1,44	0,09
HC633b	1,45	0,07
HC639a	1,50	0,27
HC639b	1,54	0,28
HC641a	1,42	0,07
HC641b	1,42	0,14
HC642a	1,80	0,51
HC642b	1,88	0,58
HC644a	1,87	0,49
HC644b	1,65	0,33
HC656a	1,51	0,18
HC656b	1,44	0,09
HC657a	1,40	0,04
HC657b	1,53	0,26
HC658a	1,70	0,34
HC658b	1,70	0,42

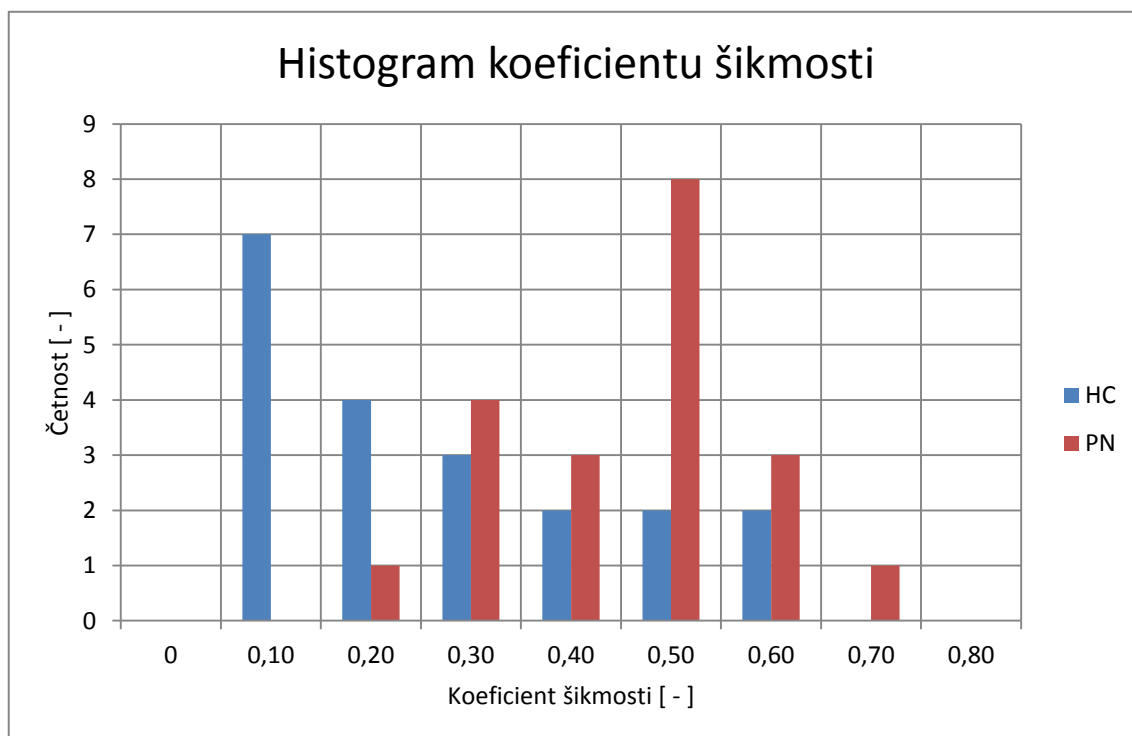
Tabulka 5.1 Zjištěné hodnoty pro HC

Subjekt	Špičatost	Šikmost
PN01a	1,81	0,56
PN01b	1,95	0,69
PN10a	1,84	0,57
PN10b	1,66	0,47
PN12a	1,53	0,26
PN12b	1,60	0,37
PN13a	1,73	0,42
PN13b	1,83	0,48
PN14a	1,85	0,46
PN14b	1,87	0,57
PN15a	1,55	0,26
PN15b	1,79	0,49
PN19a	1,48	0,18
PN21a	1,84	0,44
PN21b	1,61	0,21
PN413a	1,74	0,48
PN420a	1,53	0,38
PN420b	1,60	0,45
PN501a	1,52	0,33
PN501b	1,47	0,28

Tabulka 5.2 Zjištěné hodnoty pro PN



Obrázek 5.1 Histogram koeficientu špičatosti průměrného hlasivkového pulzu pro HC i PN.



Obrázek 5.2 Histogram koeficientu šikmosti průměrného hlasivkového pulzu pro HC i PN.

U grafu histogramu koeficientu špičatosti nebylo možné pozorovat výraznější

změny špičatosti mezi lidmi z kontrolní skupiny a lidmi s diagnostikovanou Parkinsonovou nemocí.

U grafu histogramu koeficientu šikmosti je možné pozorovat, že u lidí s diagnostikovanou Parkinsonovou nemocí nabývá šikmost vyšších hodnot než u jedinců ze zdravé kontrolní skupiny. Díky tomuto zjištění vyplívá, že by koeficient šikmost pulzu mohl být využit jako jeden z příznaků pro detekování Parkinsonovi nemoci.

ZÁVĚR

První kapitola se dělí na několik menších podkapitol. Nejdříve se tato práce věnuje popisu vzniku mluvené řeči a jejím vlastnostem, je zde popsán celý hlasový trakt skládající se z několika součinně pracujících orgánů. Následuje podkapitola zabývající se zběžným popisem metod pro přístrojové měření tvaru hlasivkových pulzů. Na tu je volně navázáno s podkapitolou obsahující přehled a popis softwarových nástrojů umožňujících určení hlasivkových pulzů z řečového signálu. Poslední podkapitola nastiňuje možnosti využití získaných hlasivkových pulzů.

Druhá kapitola obsahuje kompletní popis metod použitých v této práci pro odhad tvaru hlasivkového pulzu z řečového signálu. V této kapitole je popis pouze „hlavního“ algoritmu pro odhad pulzů metodou IAIF a Sahoo.

Ve třetí kapitole jsou ukázány výstupy naprogramovaných metod a porovnány s výstupem z programu Aparat.

Čtvrtá kapitola se věnuje popisu vytvořeného grafického uživatelského prostředí (GUI) v prostředí MATLAB 2015a a je rozdělena na menší kapitoly. První menší kapitola se věnuje popisu jednotlivých bloků GUI, kde každý bloku má příslušnou jednu podkapitolu. V těchto podkapitolách je vždy stručný popis použitých tlačítek, přepínačů a indikátorů pro jednotlivé popisované bloky. Následující menší kapitola byla ve stejném stylu jako předchozí rozdělena na podkapitoly věnující se jednotlivým blokům, ovšem s tím rozdílem že jsou zde podrobnější informace o jednotlivých tlačítkách (přepínačích) a také ukázky co se stane po jejich stisku.

V páté kapitole je provedeno srovnání tvaru hlasivkových pulzů zdravých lidí a jedinců s diagnostikovanou Parkinsonovou nemocí. Byly sledovány koeficienty špičatosti a šikmosti hlasivkových pulzů. Ze zjištěných dat bylo vyzorováno, že koeficient šikmost pulzu by mohl být využit jako jeden z příznaků pro detekování Parkinsonovi nemoci z řečového signálu.

Zhodnocení metody IAIF, její výhodou je rychlost zpracování řeči a následný odhad hlasivkových pulzů, dokáže odhadnout pulzy i v signálu s větším poměrem S/N. Je vhodná pro analyzování krátkých, tak i dlouhých úseků řeči.

Zhodnocení metody Sahoo, výhodou je že metoda pracuje se skutečným (válcovým) modelem hlasového traktu a přizpůsobuje vypočtená data na derivovaný LF-model hlasivkového pulzu. Protože rozložení hlasového traktu je jiné pro každou samohlásku, není tato metoda vhodná pro analyzování dlouhých úseků řeči z důvodu nutnosti měnit koeficienty hlasového traktu.

LITERATURA

- [1] PSUTKA, Josef. *Mluvíme s počítačem česky*. Praha: Academia, 2006. Česká matice technická (Academia). ISBN 80-200-1309-1.
- [2] HOW ELECTROGLOTTOGRAPHY WORKS. In: *Electroglottography* [online]. [cit. 2017-12-04]. Dostupné z: <http://www.rose-medical.com/electroglottography.html>
- [3] TITZE, Ingo R., Brad H. STORY, Gregory C. BURNETT, John F. HOLZRICHTER, Lawrence C. NG a Wayne A. LEA. Comparison between electroglottography and electromagnetic glottography. *The Journal of the Acoustical Society of America* [online]. 2000, 107(1), 581-588 [cit. 2017-12-04]. DOI: 10.1121/1.428324. ISSN 0001-4966. Dostupné z: <http://asa.scitation.org/doi/10.1121/1.428324>
- [4] Glottal. *Glottal* [online]. Syracuse, NY: Glottal Enterprises, 2015 [cit. 2017-12-08]. Dostupné z: <http://www.glottal.com/Aeroview.html>
- [5] Laryngeal Videostroboscopy. *Chicago Institute For Voice Care* [online]. Chicago: Chicago Institute For Voice Care, .. [cit. 2017-12-06]. Dostupné z: <http://www.chicagovoicecare.com/videostroboscopy.html>
- [6] Assessment – Vocal Nodules. In: *Vocal Nodules* [online]. [cit. 2017-12-04]. Dostupné z: <https://wkuvocalnodules.weebly.com/assessment.html>
- [7] Vyšetřovací metody hlasu. *Základy otorinolaryngologie a foniatrie* [online]. Brno: Klinika dětské ORL Lékařské fakulty MU a Fakultní Nemocnice Brno, 2009 [cit. 2017-12-06]. Dostupné z: https://is.muni.cz/elportal/estud/pdf/js09/orl/web/pages/7_2_vysetrovaci_metody_hlasu.html
- [8] TSUJI, Domingos, Adriana HACHIYA, Maria DAJER, Camila ISHIKAWA, Marystella TAKAHASHI a Arlindo MONTAGNOLI. Improvement of Vocal Pathologies Diagnosis Using High-Speed Videolaryngoscopy. *International Archives of Otorhinolaryngology* [online]. 2014, 18(03), 294-302 [cit. 2017-12-06]. DOI: 10.1055/s-0034-1372512. ISSN 1809-9777. Dostupné z: <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0034-1372512>
- [9] *Aalto Aparat Manual v2.0* [online]. Aalto University School of Electrical Engineering, 2015 [cit. 2017-12-09]. Dostupné z: <http://research.spa.aalto.fi/projects/aparat/AaltoAparatManual.pdf> . Manual. Aalto University School of Electrical Engineering.
- [10] Sopran. *Tolvan Data* [online]. Stockholm: Tolvan Data, c2017 [cit. 2017-12-09]. Dostupné z: <http://www.tolvan.com/index.php?page=/sopran/sopran.php>
- [11] *Analysis and Synthesis of Pathological Voice Quality* [online]. Los Angeles, 2016 [cit. 2017-12-09]. Dostupné z: <https://www.uclahealth.org/head-neck-surgery/workfiles/Laryngeal%20Voice%20Research/Software/Analysis%20and%20Synthesis%20of%20Pathological%20Voice%20Quality.pdf> . Manual. The University of California.
- [12] JAVKIN, Hector Raul, Norma ANTON~ANZAS-BARROSO a Ian MADDIESON. Digital Inverse Filtering for Linguistic Research. *Journal of Speech Language and Hearing Research* [online]. 1987, 30(1), 122- [cit. 2017-12-09]. DOI: 10.1044/jshr.3001.122. ISSN 1092-4388. Dostupné z: <http://jslhr.pubs.asha.org/article.aspx?doi=10.1044/jshr.3001.122>
- [13] ROSENBERG, A. E. Effect of Glottal Pulse Shape on the Quality of Natural Vowels. *The*

- Journal of the Acoustical Society of America*[online]. 1971, **49**(2B), 583-590 [cit. 2017-12-09]. DOI: 10.1121/1.1912389. ISSN 0001-4966. Dostupné z: <http://asa.scitation.org/doi/10.1121/1.1912389>
- [14] A four-parameter model of glottal flow. *STL-QPSR* [online]. 1985, **26**(4), 001-013 [cit. 2017-12-09]. Dostupné z: <https://www.semanticscholar.org/paper/A-four-parameter-model-of-glottal-flow-Liljencrants-Lin/d8e775f5b141505d38f04def2c80a583008642ac?tab=abstract>
- [15] FUJISAKI, H. a M. LJUNGQVIST. Proposal and evaluation of models for the glottal source waveform. In: *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing* [online]. Institute of Electrical and Electronics Engineers, 1986, s. 1605-1608 [cit. 2017-12-09]. DOI: 10.1109/ICASSP.1986.1169239. Dostupné z: <http://ieeexplore.ieee.org/document/1169239/>
- [16] VELDHUIS, Raymond. A computationally efficient alternative for the Liljencrants–Fant model and its perceptual evaluation. *The Journal of the Acoustical Society of America* [online]. 1998, **103**(1), 566-571 [cit. 2017-12-09]. DOI: 10.1121/1.421103. ISSN 0001-4966. Dostupné z: <http://asa.scitation.org/doi/10.1121/1.421103>
- [17] DRUGMAN, Thomas, Paavo ALKU, Abeer ALWAN a Bayya YEGNANARAYANA. Glottal source processing: From analysis to applications. *Computer Speech & Language* [online]. 2014, **28**(5), 1117-1138 [cit. 2017-12-09]. DOI: 10.1016/j.csl.2014.03.003. ISSN 08852308. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0885230814000229>
- [18] ZEN, Heiga, Keiichi TOKUDA a Alan W. BLACK. Statistical parametric speech synthesis. *Speech Communication* [online]. 2009, **51**(11), 1039-1064 [cit. 2017-12-10]. DOI: 10.1016/j.specom.2009.04.004. ISSN 01676393. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0167639309000648>
- [19] CABRAL, J.P, Renals,S., Richmond, K., Yamagishi, J.: Glottal Spectra Separation for Parametric Speech Synthesis. *Proceedings Interspeech* [online]. 2008, 1829-1832 [cit. 2017-12-10]. Dostupné z: <https://www.era.lib.ed.ac.uk/handle/1842/3832>
- [20] DEGOTTEX, Gilles, Axel ROEBEL a Xavier RODET. Pitch transposition and breathiness modification using a glottal source model and its adapted vocal-tract filter. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* [online]. IEEE, 2011, s. 5128-5131 [cit. 2017-12-10]. DOI: 10.1109/ICASSP.2011.5947511. ISBN 978-1-4577-0538-0. Dostupné z: <http://ieeexplore.ieee.org/document/5947511/>
- [21] REYNOLDS, Douglas A. An overview of automatic speaker recognition technology. In: *IEEE International Conference on Acoustics Speech and Signal Processing* [online]. IEEE, 2002, IV-4072-IV-4075 [cit. 2017-12-10]. DOI: 10.1109/ICASSP.2002.5745552. ISBN 0-7803-7402-9. Dostupné z: <http://ieeexplore.ieee.org/document/5745552/>
- [22] DRUGMAN, Thomas a Thierry DUTOIT. The Deterministic Plus Stochastic Model of the Residual Signal and Its Applications. *IEEE Transactions on Audio, Speech, and Language Processing* [online]. 2012, **20**(3), 968-981 [cit. 2017-12-10]. DOI: 10.1109/TASL.2011.2169787. ISSN 1558-7916. Dostupné z: <http://ieeexplore.ieee.org/document/6029971/>
- [23] KINNUNEN, Tomi a Paavo ALKU. On separating glottal source and vocal tract information in telephony speaker verification. In: *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* [online]. IEEE, 2009, s. 4545-4548 [cit. 2017-12-10]. DOI: 10.1109/ICASSP.2009.4960641. ISBN 978-1-4244-2353-8. Dostupné z: <http://ieeexplore.ieee.org/document/4960641/>

- [24] OZDAS, A., R.G. SHIAVI, S.E. SILVERMAN, M.K. SILVERMAN a D.M. WILKES. Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk. *IEEE Transactions on Biomedical Engineering* [online]. 2004, **51**(9), 1530-1540 [cit. 2017-12-10]. DOI: 10.1109/TBME.2004.827544. ISSN 0018-9294. Dostupné z: <http://ieeexplore.ieee.org/document/1325813/>
- [25] QUATIERI, Thomas F. ; Malyska, Nicolas: Vocal-Source Biomarkers for Depression: A Link to Psychomotor Activity.. In: *INTERSPEECH* : ISCA, 2012, S. 1059-1062 [cit. 2017-12-10]. Dostupné z: https://www.ll.mit.edu/mission/cybersec/publications/publication-files/full_papers/2012_09_09_MalyskaN_Interspeech_FP.pdf
- [26] BELALCÁZAR-BOLAÑOS, Elkyn Alexander, Juan Rafael OROZCO-ARROYAVE, Jesús Francisco VARGAS-BONILLA, Tino HADERLEIN a Elmar NÖTH. Glottal Flow Patterns Analyses for Parkinson's Disease Detection: Acoustic and Nonlinear Approaches. SOJKA, Petr, Aleš HORÁK, Ivan KOPEČEK a Karel PALA, ed. Text, Speech, and Dialogue [online]. Cham: Springer International Publishing, 2016, s. 400-407 [cit. 2017-12-11]. Lecture Notes in Computer Science. DOI: 10.1007/978-3-319-45510-5_46. ISBN 978-3-319-45509-9. Dostupné z: http://link.springer.com/10.1007/978-3-319-45510-5_46
- [27] ALKU, Paavo. Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering. *Speech Communication* [online]. 1992, **11**(2-3), 109-118 [cit. 2017-12-10]. DOI: 10.1016/0167-6393(92)90005-R. ISSN 01676393. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/016763939290005R>
- [28] PULLAKA, Hannu. Analysis of Human Voice Production Using Inverse Filtering, High-Speed Imaging, and Electroglottography [online]. Espoo, 2005 [cit. 2017-12-10]. Dostupné z: <https://aaltodoc.aalto.fi/handle/123456789/982>. Master's Thesis. HELSINKI UNIVERSITY OF TECHNOLOGY.
- [29] EL-JAROUDI, A. a J. MAKHOUL. Discrete all-pole modeling. *IEEE Transactions on Signal Processing* [online]. **39**(2), 411-423 [cit. 2017-12-10]. DOI: 10.1109/78.80824. ISSN 1053587x. Dostupné z: <http://ieeexplore.ieee.org/document/80824/>
- [30] SAHOO, Subhasmita a Aurobinda ROUTRAY. A Novel Method of Glottal Inverse Filtering. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* [online]. 2016, **24**(7), 1230-1241 [cit. 2017-12-13]. DOI: 10.1109/TASLP.2016.2551864. ISSN 2329-9290. Dostupné z: <http://ieeexplore.ieee.org/document/7448892/>
- [31] STORY, B. H., I. R. Titze, a E. A. Hoffman. Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Amer.*, [online] **100**(1), 537-554 [cit. 2017-12-10]. Dostupné z: <http://sal-slhs.webhost.uits.arizona.edu/sites/default/files/STH1996.pdf>
- [32] PLUMPE, M.D., T.F. QUATIERI a D.A. REYNOLDS. Modeling of the glottal flow derivative waveform with application to speaker identification. *IEEE Transactions on Speech and Audio Processing*[online]. **7**(5), 569-586 [cit. 2018-05-12]. DOI: 10.1109/89.784109. ISSN 10636676. Dostupné z: <http://ieeexplore.ieee.org/document/784109/>
- [33] *Modeling the Vocal Tract Transfer Function Using a 3D Digital Waveguide Mesh* [online]. University of York, UK: Audio Lab, Department of Electronics, 2012 [cit. 2018-05-10]. Dostupné z: <http://www-users.york.ac.uk/~dtm3/3DVTTF/index.html>
- [34] SIGMUND, M. *Rozpoznávání řečových signálů*. Skriptum FEKT VUT v Brně. Brno: MJ servis, 2007.
- [35] *BIOSIGNAL CHALLENGE 2018: Soubor zvukových nahrávek* [online]. Elektrotechnická fakulta ČVUT v Praze, Neurologická klinika 1. LF UK, HUMUSOFT a MathWorks, 2018 [cit. 2018-05-16]. Dostupné z: <http://sami.fel.cvut.cz/biosignalchallenge2018/>

SEZNAM PŘÍLOH

A	Popis algoritmu pro metodu discrete all-pole modelling (DAP) [29]	58
B	Algoritmus pro určení fáze uzavření hlasivek z řečového signálu [32]	59
C	Metody pro nalezení znělých úseků z řečového signálu	62
C.1	Metoda krátkodobé intenzity [1]:	62
C.2	Metoda krátkodobé funkce středního počtu průchodů signálu nulou [1]:	62
C.3	Metoda základního tónu řeči za použití Center-Clipping [34]:	62

SEZNAM SOUBORŮ NA PŘILOŽENÉM CD

- Elektronická verze textu diplomové práce
- Program na získání hlasivkových pulzů z průběhu řečového signálu
- Vzorové řečové signály

A POPIS ALGORITMU PRO METODU DISCRETE ALL-POLE MODELLING (DAP) [29]

- 1) Pomocí rychlé Fourierovy transformace (FFT) se dostaneme z časové oblasti signálu do frekvenční oblasti. Zde detekované špičky signálu odpovídající formantovým frekvencím (u znělých částí), zjistíme jejich pozici ω_m a výkonovou hodnotu $P(\omega_m)$.
- 2) Následné je vypočítána autokorelace [29]:

$$R(i) = \frac{1}{N} \sum_{m=1}^N P(\omega_m) e^{j\omega_m i}, \quad (\text{A.1})$$

kde i náleží do intervalu $1 \leq i \leq p$ a p je požadovaný řád filtru, m náleží do intervalu $1 \leq m \leq N$ a N je počet nalezených špiček (formantů) ve spektru.

- 3) Použít LPC pro počáteční odhad koeficientů a .
- 4) Vypočítat [29]:

$$A(\omega_m) = \sum_{k=0}^p a_k e^{-j\omega_m k}, \quad (\text{A.2})$$

m náleží do intervalu $1 \leq m \leq N$.

- 5) Vypočítat časově obrácenou impulzní odezvu [29]:

$$\hat{h}(-i) = \frac{1}{N} \sum_{m=1}^N \frac{e^{-j\omega_m i}}{A(\omega_m)}, \quad (\text{A.3})$$

kde i náleží do intervalu $0 \leq i \leq p$.

- 6) Vypočítat nový parametr pro a [29]:

$$a_{m+1} = a_m(1 - \alpha) + \alpha R^{-1} \hat{h}_m, \quad (\text{A.4})$$

kde $\alpha = 0,5$ určuje míru rychlosti, s jakou probíhá konvergence.

- 7) Vypočítat míru chyby podle Itakura-Saito [29]:

$$E_{IS} = \frac{1}{N} \sum_{m=1}^N \frac{P(\omega_m)}{\hat{P}(\omega_m)} - \ln \frac{P(\omega_m)}{\hat{P}(\omega_m)} - 1, \quad (\text{A.5})$$

kde $\hat{P}(\omega_m)$ je:

$$\hat{P}(\omega_m) = \frac{1}{\left| \sum_{k=0}^p a_k e^{-j\omega_m k} \right|^2}, \quad (\text{A.6})$$

při výpočtu $\hat{P}(\omega_m)$ se za hodnotu a_k dosadí výsledek z bodu č. 6. Pokud je

výsledek chyby E_{IS} vyšší než zvolený práh vrací se algoritmus zpět k bodu č. 4, dokud není dosaženo požadovaného prahu.

8) Znormování dosažených hodnot podle [29]:

$$\frac{1}{N} \sum_{m=1}^N \frac{P(\omega_m)}{\bar{P}(\omega_m)} = 1. \quad (\text{A.7})$$

9) Konec algoritmu.

B ALGORITMUS PRO URČENÍ FÁZE UZAVŘENÍ HLASIVEK Z ŘEČOVÉHO SIGNÁLU [32]

Fáze uzavření hlasivkové štěrbiny je nalezena v časovém úseku znělé řeči, během této doby je formantová modulace stacionární.

- 1) Data jsou zpracovávány po úsecích délky dvou period základního kmitočtu hlasu s překrytím odpovídající délce jedné základní periody. Na takto segmentované části řečového průběhu je aplikován „whitening“ pomocí inverzní filtrace. Koeficienty pro inverzní filtr jsou získány metodou LPC řádu $p = 14$.
- 2) Nalezení špiček pro jednotlivé úseky.
- 3) Vybrání části řeči náležící do časového okamžiku mezi současnou a následující špičkou.
- 4) Nastavení velikosti analyzačního okna podle [32]:

$$N_w = N/4, \quad (\text{B.1})$$

kde N je počet vzorků mezi dvěma špičkami a náleží do intervalu $p + 3 \leq N_w \leq 2p$.

$$N_w = p + 3, \quad (\text{B.2})$$

pokud je výsledek z rovnice (B.1) v intervalu $N_w \leq p + 3$.

$$N_w = 2 * p, \quad (\text{B.3})$$

pokud je výsledek z rovnice (B.1) v intervalu $N_w \geq 2 * p$.

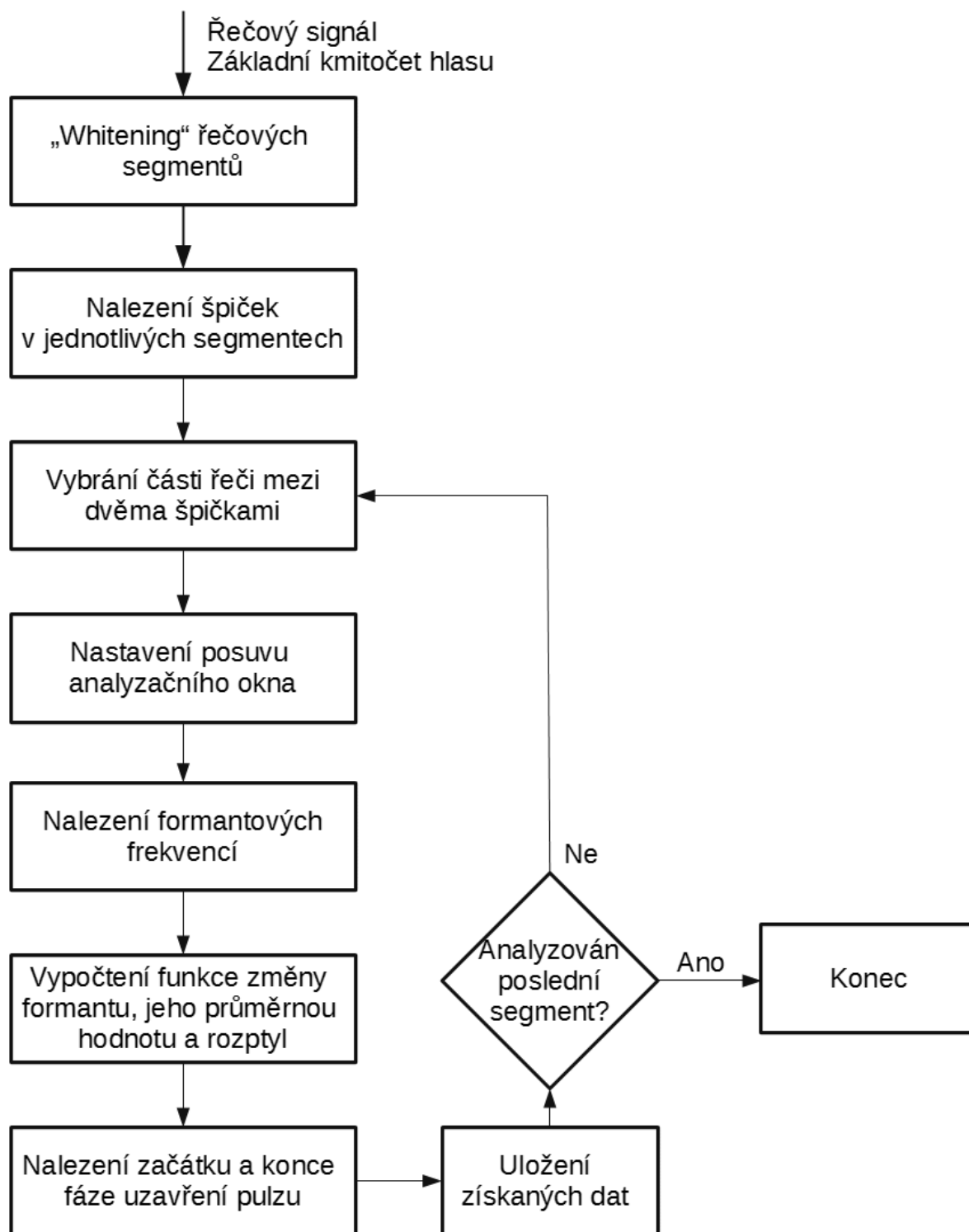
- 5) Pro nalezení formantových frekvencí je použito analyzační okno o délce N_w . To se postupně posouvá po úseku vybrané části řeči s krokem jednoho vzorku (od začátku po konec úseku), při každém posunu jsou nalezeny nové hodnoty formantových frekvencí.

- 6) Ze zjištěných hodnot formantů je vypočítána funkce popisující změnu první formantové frekvence [32]:

$$D(n_0) = \sum_{i=n_0}^{n_0+4} |F(i) - F(i-1)|, \quad (\text{B.4})$$

kde n_0 náleží do intervalu $1 \leq n_0 \leq N - N_w - 5$, N je počet vzorků ve vybraném úseku a F je hodnota první formantové frekvence. Při nalezení minimální hodnoty $D(n_0)$ je určen počátek „stacionární formantové oblasti“, ve které se zjistí průměrná velikost a směrodatná odchylka prvního formantu pro tuto oblast.

- 7) Stacionární oblast může být rozšířena na **pravou stranu**, pokud je rozdíl následujícího a průměrného formantu menší než dvě směrodatné odchylky. Po rozšíření oblasti je nutné přepočítat velikost a směrodatnou odchylku prvního formantu. Stacionární oblast může být rozšířena na **levou stranu**, pokud je rozdíl předcházejícího a průměrného formantu menší než dvě směrodatné odchylky. Průměrná velikost a směrodatná odchylka se nepřepočítává, zůstává (stejná) na konečné hodnotě zjištěné při rozšiřování na pravou stranu.
- 8) Uložení získaných dat informujících o začátku a konci stacionárního úseku.
- 9) Pokud nebyl analyzován poslední segment mezi dvěma špičkami, pokračuje se znovu od bodu č. 3.



Obrázek B.1 Algoritmus pro určení fáze uzavření hlasivek z řečového signálu.

C METODY PRO NALEZENÍ ZNĚLÝCH ÚSEKŮ Z ŘEČOVÉHO SIGNÁLU

Pro nalezení znělých úseků v řeči bylo využito tří metod. Jako znělý úsek je označena část signálu, při kterém byla překročena zvolená prahová úroveň použité metody.

C.1 Metoda krátkodobé intenzity [1]:

$$M_n = \sum_{k=-\infty}^{\infty} |s(k)| w(n - k), \quad (\text{C.1})$$

kde $s(k)$ je vzorek signálu v čase k a $w(n)$ udává typ použitého okénka (Hammingovo).

C.2 Metoda krátkodobé funkce středního počtu průchodů signálu nulou [1]:

$$Z_n = \sum_{k=-\infty}^{\infty} |\text{sgn}[s(k)] - \text{sgn}[s(k - 1)]| w(n - k), \quad (\text{C.2})$$

kde $\text{sgn}[s(k)]$ je:

$$\text{sgn}[s(k)] = \begin{cases} 1 & \text{pro } s(k) \geq 0 \\ -1 & \text{pro } s(k) < 0 \end{cases} \quad (\text{C.3})$$

kde $w(n)$ udává typ použitého okénka (pravoúhle).

C.3 Metoda základního tónu řeči za použití Center-Clipping [34]:

- 1) Signál se rozdělí na segment o délce 10 ms.
- 2) Určení prahové hodnoty Clipping Level (C_L) pro zpracováváný segment podle [32]:

$$C_{Lj} = k \min(\text{MAX}_{j-1}, \text{MAX}_{j+1}), \quad (\text{C.4})$$

kde k je redukční faktor, MAX_{j-1} je maximum z předchozího segmentu a MAX_{j+1} je maximum z nadcházejícího segmentu.

- 3) Převod řečového signálu $s(n)$ na posloupnost $\tilde{s}(n)$ podle [34]:

$$\tilde{s}(n) = \begin{cases} +1 & \text{pro } s(n) > C_L \\ 0 & \text{pro } -C_L \leq s(n) \leq C_L \\ -1 & \text{pro } s(n) < -C_L \end{cases} \quad (\text{C.5})$$

4) Výpočet autokorelační funkce signálu $\tilde{s}(n)$ podle [34]:

$$\tilde{R}(k) = \sum_{n=1}^{N-k} \tilde{s}(n)\tilde{s}(n+k), \quad (\text{C.6})$$

kde $k = 0, 1, \dots, N-1$.

5) Úsek je znělý pokud platí podmínka [34]:

$$\tilde{R}(k)_{\max} < \alpha \tilde{R}(0), \quad (\text{C.7})$$

kde $\tilde{R}(k)_{\max}$ je hodnota prvního maxima a $\alpha (\approx 0,3)$ slouží k nastavení prahové úrovně.